Industrial Management Review

Fall 1961

School of Industrial
Management
Massachusetts Institute
of Technology

The Industrial Management Review is published twice each year by the Industrial Management Review Association, Massachusetts Institute of Technology, 50 Memorial Drive, Cambridge 39.

Editors John V. Olszewski, Michael Roemer, and Fred I. Steele.

Subscriptions $2.00 per year (two issues); 3 years $5.00. Reprints are available upon request.

Copyright 1961 by the Industrial Management Review Association.

Manuscripts should be submitted to the Editors, Industrial Management Review, Room 52–472.

Industrial Management Review

Vol. 3, No. 1

School of Industrial Management, Massachusetts Institute of Technology

Fall, 1961

Contents

1 The Manager and His World — A. T. M. Wilson

27 Captive Finance Companies: Their Growth and Some Speculations on Their Significance — Victor L. Andrews

56 Management Decision Making: Some Research — Edward H. Bowman

64 Organization Structure: Social and Technical Elements — Chadwick J. Haberstroh

78 An Approach to Job Shop Scheduling Using a Large-Scale Computer — James C. Emery

A. T. M. Wilson
Unilever, Ltd.

The Manager and His World

The first section of this paper, "The Boundary Position of the Manager," describes the pattern of activities, largely external to the firm, with which the high level manager is concerned. These activities may be considered as transactions which take place across the boundary between the firm and various groups with which it is related - suppliers and consumers, for example. The discussion is based on a number of general concepts or theories of the firm, selected to illustrate the convergence of approach which is becoming apparent in studies of the firm by writers on management[6, 7, 9, 13], by business economists[2], by sociologists[14] or by those concerned with applications of system theory[3, 15]. I must ask my readers to tolerate the risk of confusion liable to arise from parallel consideration of "the firm" and "the manager" in this discussion.

The second section, headed "Management Development as an Internal Institution of the Firm," discusses some of the problems which arise from the strength of the internal and external forces around the whole process of management development. It begins with examples of discontinuities in the career patterns of managers and goes on to use a particular component of one management development scheme - a component concerned with initial recruitment for management - to exemplify some of the institutional aspects of problems at this particular boundary of the firm and of the manager's role.

There have been three main sources for the concepts and comments which make up this contribution. In the first section the main source is that of the empirical studies and the writing of the authors quoted. The second section is largely based on the work of my former colleagues at the Tavistock Institute, a research and consultant group which operates in the social science field and, so far as industry is concerned, works mainly at the level of the firm. Finally, the views expressed have been considerably influenced by recent experience of collaboration, from within, in the

---

[1] Presented as a background paper at the Symposium on Executive Development, M.I.T., April 1961.

efforts of Unilever to grapple with its problems of executive development and to seek means of making improved use of the social sciences in this field and in others.

I. THE BOUNDARY POSITION OF THE MANAGER

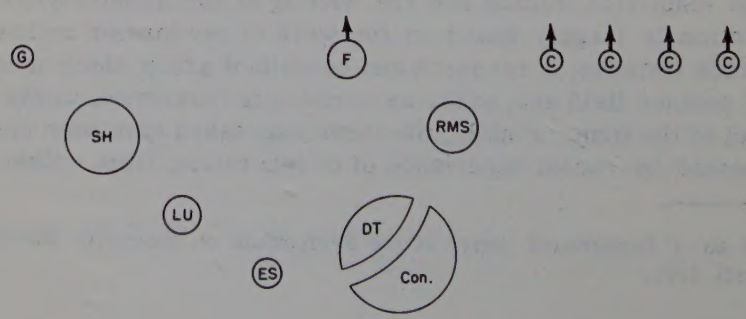
Despite differences arising from the varying size and character of firms, the main tasks of all higher level managers have a considerable resemblance. This section concerns one way in which the common elements in these tasks can be considered - that is, in terms of the pattern of relations and transactions between the manager and eight areas of relevant social activity, most of which contain a number of significant social institutions and social groups. These areas may be rather summarily labelled as:

1. Government;
2. Distributors and consumers;
3. Shareholders;
4. Competitors;
5. Raw material and power suppliers;
6. Sources of employees, particularly managers;
7. Trades Unions;
8. Groups within the firm.

All of these - and the manager himself - form part of a single community with a pervading general influence, and many individuals and groups would appear, in different aspects, in more than one of the areas listed. But for present purposes the differentiation of these eight areas may be adequate for the general points I want to introduce and to illustrate.

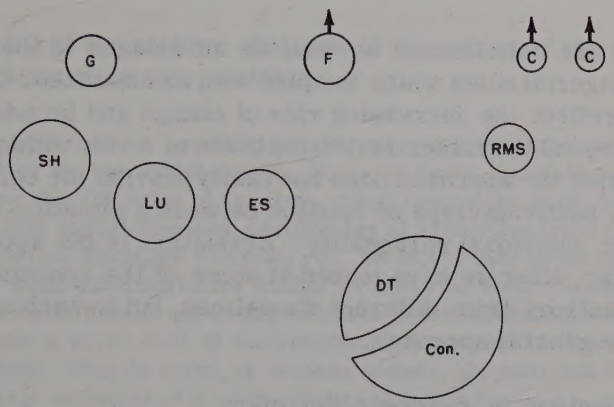
The manager's central tasks may be described as those of formulating agreed policies and plans which will maintain movement towards the general objectives of his firm, and which will, indeed, continuously redefine and pursue these objectives. For these purposes, he needs to build and maintain an appropriate type and pattern of relationships and transactions between himself and others within his firm and, externally, with the relevant groups and individuals in each of the areas listed.

Diagram I - 1925-1940



F is the firm; G is government; SH is shareholders; LU is Labor Unions; C is competitor; RMS is raw material supplier; DT and Con. are distributors and consumers; ES is employment sources. The size of the circles and their distance from that representing the firm represent to some degree the extent and the intimacy of the influences which each external group has on the firm. The differences between Diagrams I and II give some indication - inevitably inaccurate - of familiar trends over the past twenty years.

Diagram II - 1960



The main points of interest are represented in Diagrams I and II. These are intended in particular to illustrate the increasing modal size of the firm; the much increased influence of government, of labor unions, and of employee sources - particularly potential managers; and, most obvious of all, the greatly increased power of the consumer and of the distributors[5, 8, 11]. The power of shareholders has diminished and - despite the legal situation - they are shown as an "external" group. Allied to these changes is the emergence of the "professional" as against the owner manager, and perhaps most important of all, the increasing mutual proximity of the external areas with which the manager must maintain an appropriate pattern of relationships and transactions. Attempts to construct these diagrams so that they directly illustrated the "boundary position" of the manager seemed to make them excessively complex, and, despite the distances between the circles, the higher-level manager needs to be in contact, in varying degrees of directness, with each of the external areas represented. The words "boundary position" refer not only to certain theoretical approaches derived from Lewinian psychology and system theory, but also to the practical point that the manager must strive to achieve an optimal degree of "boundary control" in relation to his transactions across each of these boundaries if he is to be successful in providing for the adaptation and growth of his firm.

To some extent, of course, every type of job, every role, in every type of organization and institution, has a general character of this kind. But what is particularly significant for our discussion is to consider how far the role of a manager in industry calls for a distinctive competence, and if it does, how this can best be described. It seems clear, for example, that any distinctive character in a manager's role will be linked to the tasks of maintaining progress towards objectives while handling problems of change in each of the main areas with which he must deal. It must certainly take into account skill in dealing with the increasing speed of change and with the existence of awkward inconsistencies and conflicts in the pattern of relations which must be built and maintained.

There are deficiencies as well as advantages in the general approach to managerial roles which has just been summarized. One advantage is that it can reflect the increasing rate of change and broaden discussion of managers beyond the rather restricted basis of a role within a firm. One deficiency is that the approach does not easily provide for consideration of leadership - a particular type of relation between a situation, a role within a group and an individual personality. Evaluation of the approach will be easier, however, after we have looked at some of the concepts of the firm developed by authors from different disciplines, but in each case exemplifying the same general approach.

Business in Society: a European Summary

Two years ago, at a conference organized by the Organization for European Economic Cooperation, Professor Charles Celier[4] of the Institute of Political Studies at the Sorbonne, summarized the impact on management of developments in Western Europe:

> It would be no paradox to state that the importance assumed by political, economic and social phenomena in questions of management is one of the most noteworthy consequences of the general development of Western Europe in the last twenty years. Looking back to the eve of the Second World War, it was evident that many firms could then carry on their activities without usually worrying about the general situation of the society in which they existed. This is no longer true . . . the modern firm needs, more and more, strictly defined specialization; it has also to call upon an increasing number of outside services. More often than not, it does not even work on raw material but on semi-finished products . . . Its own rate of work is governed by that of its sub-contractors. The function of management is thus to coordinate large numbers of different activities, many of which are outside its direct authority, as they derive from outside sources. It has thus become impossible to exercise such a paternal authority as was formerly the custom. The head of a firm cannot be skilled in all the techniques he needs to employ; he cannot set himself up as a model to all the different specialized and highly skilled trained assistants he requires . . .

Celier goes on to illustrate a point which is possibly less controversial than it was - patrimonial and paternalist management, by its nature, tends to inhibit recognition of important changes and may even prevent adequate communication:

> Many French manufacturers who traditionally maintained relations of mutual trust with their workers thought they knew the staff attached to their firm and felt perfectly capable of foreseeing all their reactions . . . (They) were amazed when, in 1936, they saw their workers occupying their factories in obedience to the orders of a hitherto unheard of trade union.

He goes on to fill in the picture of change in the nature of the relationships which the manager must now maintain:

> The manager has to negotiate almost every day with administrative authorities in the strict sense. Some of these maintain constant watch on the activities of the firm, which frequently needs permits, import and export licenses, certificates of exemption from a social or health law, exchange controls, etc. He has to turn to administrative or semi-administrative bodies in order to obtain credit, take out insurance, branch out in new markets under adequate guarantees, or collect a great deal of information only obtainable from public authorities. What is more, in modern society, the state has become the biggest customer of a large number of industries . . .

After discussing such international problems as currency and purchasing power, Celier concludes:

> It is thus almost impossible to separate economic factors, properly speaking, from the social and political factors which also play a large part in the life of a firm . . . heads of industrial and commercial firms (have a role which) is centered quite as much on measures which will enable the firm to adapt itself successfully to the changing face of the . . . surroundings in which it pursues its activities as on solving internal technical or commercial problems which can now be delegated to skilled specialists. The function of the head of the firm is constantly to adapt the unit which he governs to the surroundings from which it draws its sustenance and in which it has to live and prosper. This quite novel conception would certainly surprise the pioneers, who, some hundred or hundred and fifty years ago, established industry and finance in Western Europe. It is nevertheless a fundamental fact for the training of management staff for its future work.

Entirely similar conceptions of the enterprise emerge from other analyses made from separate starting points. These too stress the need of the firm actively to adapt to its environment in directions which are essential or desirable for its survival and growth. Approaches of this kind are dynamic, in the sense of being concerned with forces, with equilibria,

and with processes of change, and it seems clear that they describe such essential characteristics of any firm as to suggest the limits of any purely descriptive approach or any attempt to understand the firm merely in terms of its internal structure at a particular point in time. The same comments may well be made in considering management roles.

Some Definitions of "Business" and "Enterprise"

The quotations now to be given and discussed suggest types of fundamental questions on which the manager of the future will need to have considered views. And in the larger firms these views will need to provide a basis for action at the international level, as well as that of the community and the national society. Here is Peter Drucker[6]:

> The first conclusion . . . is that a business enterprise is created and managed by people. It is not managed by forces. Economic forces set limits to what management can do. They create opportunities for management's action; but they by themselves do not determine what a business is or what it does . . . The second conclusion is that a business cannot be defined or explained in terms of profit . . . This does not mean that profit and profitability are unimportant . . . Profit is not the explanation, cause, or rationale of business behavior or business decisions but the test of their validity . . . If we want to know what a business is we have to start with its purpose. And its purpose lies outside of the business itself. In fact, it must lie in society since a business enterprise is an organ of society. There is only one valid definition of business purpose: to create a customer . . . It is the customer who determines what a business is (and because of this) any business enterprise has two - and only these two - basic functions: marketing and innovation. They are the entrepreneurial functions. Marketing is the distinguishing, the unique function of the business. A business is set apart from all other human organizations by the fact that it markets a product or a service . . . (and) . . . to reach full realization of the importance of marketing requires overcoming a deep-rooted social prejudice against "selling" as ignoble and parasitical and in favor of "production" as gentlemanly, with its resultant theoretical fallacy of considering production as the main and determining feature and function of a business.
>
> The second function of business . . . innovation . . . may take the form of lower price . . . a new or better product, a new convenience, or the creation of a new want or . . . finding new uses for old products. Innovation goes right through all phases of a business . . . the most effective innovations in American history in the last few years (1950-55) were probably not the much publicized new . . . products and processes but innovations in materials handling and in manager development.

Elsewhere Drucker emphasizes that "management is not just a creature of the economy; it is a creator as well. And only to the extent

to which it masters the economic circumstances and alters them by conscious directed action does it really manage." In a slightly different context, Hurff[11] makes an equivalent point in "Social Aspects of Enterprise in a Large Corporation." "Enterprise," he writes, "is that factor in the productive process which is responsible for adjusting a business organization to a dynamic world through innovation and adaptation - a factor which historically and currently is identified with general executives and directors."

You may feel that distribution rather than production is unduly emphasized in these selected quotations, but we are primarily discussing events in societies where the changes of the past fifty years could largely be represented by this difference in the focus of attention. Indeed, another formulation of this same point would say that it may be dangerous to conceive of a firm in terms of its products rather than in terms of the changing markets, the overt consumer wants, and the latent consumer needs to which the products must be constantly geared and adapted[12].

In "The Meaning of Modern Business," one of the most recent studies in this field, Richard Eells[7] summarizes his views of the business corporation in this way:

> The modern corporation is not one thing but many things. It is not just an entrepreneurial unit, nor simply a more or less effective producer of goods and services . . . it is in part an economic institution, part educational institution, part fraternal organization, part contestant in political process.

Almost each word in this summary seems to me to have very considerable significance.

In later sections of his book, Eells goes on to consider what he calls the metrocorporation, in which,

> the "economic man" gives way to the "whole man," and we see in it the proliferation of corporate goals together with a strong tendency to over-mother the brood that clusters about this somewhat pretentious lesser society. This pretentiousness appears not only in the tendency to govern the lives of its Organization Men but also in the theory that its executive managers should seek, through policy making, to balance the interests of groups that lie both within and without the corporate orbit. In its zeal to live up to the social responsibilities increasingly thrust upon it, the metrocorporation even attempts to bring the public interest into the balance . . . As we have seen, a reaction has set in against what the traditionalists see as the overweening ambition of the metrocorporation and its allegedly dangerous departure from established norms of the profit-making function of a straight business enterprise. In other quarters, there is apprehension about corporate power and the legitimacy of author-

> ity vested in corporate officialdom . . . the search for more tenable models of corporate enterprise . . . has recently taken the form of varying kinds of doctrine about the social responsibilities of business and the businessman. It is not surprising that this should be so, for the major issues concern the external relationships of the business unit. Yet, as has been shown here, the problem involves more than external relationships and so-called public responsibilities. It also raises the relatively unexplored issue of corporate governance.

The Enterprise as an "Open" System

So far the approaches illustrated have predominantly been those of political scientists and a business economist. In those which follow the approach is of a more sociological character. Apart from their relevance as background to our theme, both the preceding quotations and those which now follow are of general interest in showing the extent to which there can be convergence in the outcome when writers from different disciplines attempt to understand the firm. Paradoxically, this convergency may facilitate distinctive approaches by contributing to a relatively solid and agreed basis from which complementary types of study can be initiated.

My former colleagues at the Tavistock Institute, Eric Trist and Fred Emery, have exemplified in a recent paper[17] their approach to the study of the business enterprise as a "socio-technical" system. They begin by drawing attention to the way in which some of the "borrowing" of sociological and psychological system theory has often been from models built up in the physical and not in the biological sciences. They stress an important distinction between the system models applicable in these two fields of traditional scientific work - a distinction clarified initially by von Bertalannfy[1] and later by Boulding[3]. The phenomena and models in the physical and in the biological sciences, Bertalannfy noted, show a considerable difference along the dimension of being relatively "closed" or relatively "open" to interplay with influences located in the immediate environment of the unity of study. Speaking about the firm, Trist and Emery point out, however, that

> In the realm of social theory there has been something of a tendency to continue thinking in terms of a "closed" system; that is, to regard the enterprise as sufficiently independent to allow most of its problems to be analyzed with reference to its internal structure and without reference to its external environment.

Such an approach - and others related to it -

> does not conceptually reflect the mutual permeation of an organization and its environment . . . Enterprises appear to possess at least these characteristics of "open systems": they grow by processes of

> internal elaboration and manage to achieve a steady state while still doing work, i.e., achieve a quasi-stationary equilibrium in which the enterprise as a whole remains constant with a continuous throughput, despite a considerable range of external changes.

Although the phrase "remains constant" needs elaboration, whether we are speaking of biological or social organisms, the general point is perhaps clear enough.

Trist and Emery go on to evaluate the concept of an "open system" as applied to business enterprises, and find that it "helps to provide a more realistic picture of how they are both influenced by and able to act back on their environments . . . arising from the nature of the enterprise as an open system, management is concerned with 'managing' both an internal system and an external environment."

They conclude that:

> The primary task of managing an enterprise as a whole is to relate the total system to its environment, and not internal regulation per se. If management is to control internal growth and development, it must in the first instance control "boundary conditions" - the forms of exchange between the enterprise and its environment . . . the strategic objective should be to place the enterprise in a position where it has some assured condition for growth . . . (and) . . . achieving this position will be the primary task of the organization.

The Concept of an Institution in Relation to the Firm

In a kind of colloquial shorthand I have become accustomed to describe an institution as a "widely agreed means to some widely agreed end," a formulation which is helpful if we wish to study the units of industry rather than industry as a whole, if we wish to consider the process of institutional development in the firm, and - the most relevant point for our immediate purpose - if we are to consider how far the growth of a firm's management development scheme towards the position of being an "internal institution" is a main criterion of its effectiveness.

Let me again quote Eells:

> The use of the term "institution" to describe the corporation is familiar enough, so that the idea of survival as a goal of business should need little qualifying. In general, that which follows a repetitive and predictable pattern is described as an institution . . . An institution, in the straight sense, therefore, is a quasi-organization which is vital enough in the whole context of human needs and human drives to be maintained by the common consent of many generations . . . It is at just such a point in organizational evolution that the corporation now finds itself.

In Leadership in Administration, Selznick[14] approaches the concept of an institution from a slightly different angle, and makes what - for our purpose - is an important distinction between an institution and an organization.

> The most striking and obvious thing about an administrative organization is . . . (that it has) . . . set aims . . . (and is) . . . conceived as an exercise in engineering; it is governed by the related ideals of rationality and discipline. The term "organization" . . . suggests a certain bareness, a lean, no-nonsense system of consciously coordinated activities. An institution, on the other hand, is more nearly a natural product of social needs and pressures - a responsive, adaptable organism. This distinction is a matter of analysis and not of direct description. The relation of an organization to the external environment is, however, only one source of institutional experience. There is also an internal social world to be considered.

In terms of lessons for management development, one general point of high relevance can be seen in these discussions of the firm as an institution. The tasks of the higher level managers center on problems in which there is a continuously high level of uncertainty, and complex value decisions are inevitably involved. This has a direct bearing on the requirements of personality for top level management, on the relevant patterns of career development, and on the objectives and content of programs of management education.[2]

In relation to the central notion we are discussing - the "open system" theory of the enterprise - great importance attaches to the transactions between a firm and its environment in the form of recruiting potential and actual managers. In the continuous adaptation of a firm and all that goes with this, it is certain that some degree of "import" and "export" of managers is almost inevitable, for personality is likely to be more resistant to change than the structure of a role or of a firm. It is indeed probable that "managerial turnover" - the exchange of managers between the firm and its environment - has an optimal level for a particular firm at a particular time. One could indeed describe the results of various kinds of ineffective top management control of transactions at this particular boundary. Inadequate control, for example, can lead either to insufficient or to excessive turnover of managers and to associated types of difficulty within the firm. A relatively high managerial turnover - there are a number of obvious varieties - can often be linked to inadequate institutional development in the firm and hence inadequate capacity to mobilize the energies and the staying power of its members;

---

[2] One difficulty in helping physical scientists to become general managers is that those trained to deal with the "closed system" phenomena of the physical sciences often find the "open system" phenomena of management - and of the social sciences - exasperatingly difficult to deal with.

a relatively low managerial turnover may be linked to a type of regressive institutionalization in which "means" have become confused with "ends" and in which the highly significant "ends" of innovation and adaptation appear low down on the scale of values held within that particular enterprise. To turn from management development to management philosophy, as it is called, there may also be ways of relating the particular stage of institutional development of a particular firm to its managerial "philosophy" and practice - for example, to an authoritarian or to a participant approach, if we use a dimension of contrast which is frequently discussed. At least in theory, it is easy to see some kind of ideological consistency in a management philosophy which regards subordinates as essentially requiring control and discipline, and which regards the external environment as predominantly offering dangerous threats rather than opportunities for creative action. These, taken together, represent a defensive approach.

II. MANAGEMENT DEVELOPMENT AS AN INTERNAL INSTITUTION OF THE FIRM

The title of this section suggests the existence of an agreed and integrated set of functions and activities which have grown to the point of being established and familiar in a considerable proportion of firms. But this situation, where it does exist, is likely to be unusual at the present time. For that reason, let me begin by saying that what is now to be discussed is the general setting in which management development schemes are initiated and some factors which influence the growth of such schemes towards the position of being an internal institution of a firm. Once more, I use the word "institution" to mean "an agreed means to some widely agreed end." Later on there is rather full discussion of some aspects of a particular scheme for external recruitment to management, and it may help to link this second section of my paper to the first, if I suggest that such a component of a complete management development scheme should really be classed as a "boundary institution" rather than as a completely "internal" institution of the firm concerned. There may well be common characteristics in all boundary institutions - but I will leave this implication to emerge.

Discontinuities in Career Patterns

Various concepts of the top manager's role have earlier been considered. If we move down the line towards the middle and lower levels of management certain important changes can be seen. Along another dimension of contrast there are equally important differences between the three main functional areas of production, marketing, and that usually labeled "administrative" or "commercial." I have been rather insistent about

common elements in the situation around top managers; for a reason shortly to be described, let me now draw attention to certain differences which can be seen between managerial groups at different levels in each of these three main functional areas - differences of considerable importance to management development as a whole.

The man at the lower end of the executive chain - he may be the supervisor, the foreman, the assistant office manager, or he may be the immediate superior of, say, eight sales representatives - has always seemed to me to be more clearly aware of the importance of the outside world and of the relevant groups around the firm than are the middle managers a little further up. Managers at first-line levels - with all the ambiguities of their position and their problem of being the representative of top management - are often in closer and more frequent contact with the outside world and with such educational and thought-provoking influences as operatives, labor union officials and distributors, all of them groups with considerable independence and all of them difficult to fob off with evasive replies to incisive questions. "It's all very well for the departmental manager," said one foreman, "to come down to my shop, sack a man and go away. He hasn't got to live afterwards with the jury of the man's mates - I have. And they know, and I know, what even short unemployment means to a family, whether or not the breadwinner was a good worker." On the marketing and sales side there is perhaps something similar in the rather cynical mutual compassion which can often be seen in the relations between the distributor, the sales representative and his manager as they transmit downwards towards the consumer the plan for a particular sales cycle and describe how successful it is going to be for everyone.

It seems as a rule to be in "middle" management that there is the most obvious problem of maintaining some continuing awareness of the world outside the immediate unit. It is at these levels that there is the greatest risk of worshipping means and forgetting ends, of over-estimating the organizational blueprint and forgetting the ineffectiveness of one-way systems of command. Little of this, I feel, can be explained in terms of the calibre of the actual men at these levels, for their intrinsic capacities seldom differ in any dramatic way from those immediately below or immediately above them. On the contrary, I feel that it is the sociological factors in their situation that matter - particularly the restricted character of the role of the middle manager. It is these factors which favor the development, at these levels, of a managerial climate which differs in rather specific ways from that of management at the top and that of the first-line level, lower down. In middle management the importance of lateral coordination, the need for close attention to timing of tactical activities, and the absence of direct contact with such relatively independent groups as operatives or consumers or distributors may well be factors which so often make this an area where management development, in the

widest sense, has its greatest difficulties, and where it cannot so far be held to have successfully held at bay a stultifying and ironically named "logic of efficiency."

It is in part for reasons of this kind that a variety of familiar institutions - industrial, military and ecclesiastical - have tended to arrange that those intended for senior positions at a later date should have no more than limited spells of experience at middle levels, and have used in planning early development what I will provocatively call an "elite" principle. "Never let your potential officer get into the sergeant's mess" was a cliche of the British regular officer between the two wars. Although even the mildest adherent of class-war theory would recognize in a moment what an expression of social prejudice that statement may contain, I want to suggest there is a great deal more in it than that. For this purpose, let me take an example which poses almost the same problem in a narrower setting - the contrast in the character of jobs at different levels within the functional area which in my own firm we call "commercial" but which elsewhere would be called "administrative." If we leave aside the changes likely to come from office mechanization and think of traditional forms of commercial work, the lowest levels in this area are largely concerned with rather routine clerical matters - carrying out and checking the whole business of invoices, bills, payments, receipts, and so on - activities which are, as a minimum, aimed at discharging statutory obligations to all those with whom money and goods or services are exchanged. A little higher up, commercial activities center on safeguarding of assets, while at the top of this commercial area there are senior managers and commercial directors whose tasks may in part require skills also needed further down - for example, a knowledge of accountancy. The point I want to make, however, is the opposite: at this top commercial level the main responsibility is the reverse of passive safeguarding of assets; it is that of trying actively to increase the assets by providing advice - and sharing decisions - about the organized gambling which makes up such an important proportion of top management work. There is good reason to believe that a man who may well have considerable potential for this high level work - it sometimes appears as a relatively late-flowering talent - may be seriously handicapped if he spends more than a short period of his career in the "asset safeguarding" lower levels where the primary responsibilities are of a quite different character. It is not uncommon, we say in England, to find that the poacher may turn gamekeeper in later life, and I hope it is not disrespectful to commercial managers to say that in their careers industry seems to attempt an opposite and intrinsically more improbable type of role reversal. There are many examples of this kind of discontinuity in career development in different functional areas in industry. All of them suggest problems of career-design within the sphere of systematic management development. They also suggest - this is my point - the need for some kind of early differentiation of men likely to become top rather than middle managers.

I want now to move away from the rather cosmic background in which the role of the manager was earlier considered, towards discussion of some equally general but rather more earthy aspects of management development. For reasons to be given in a moment, I would like to outline the bases of the types of recruitment, training and development schemes with whose initiation and growth my former Tavistock colleagues and I have been most familiar since the earliest part-model of this general kind was set up in 1941, in the form of the "War Office Selection Boards" which produced officer-cadets for training and for later commissioned rank. From this I will go on to comment on a particular recruiting component of the management development scheme in my own firm.

The Initiation of a Management Development Scheme

Let me recall once more the definition of an institution used in this paper, for I want to illustrate the thesis that management development can only be effective when it grows away from being no more than a set of procedures, and towards the position of being a "widely agreed means to a widely agreed end." Just what the limits of these ends and means are, and who must agree about them, should be defined, but I hope the general point for discussion will be clear enough.

Let me begin with some observations by American writers on management development, dealing particularly with policies and practices of recruitment and promotion. Peter Drucker[6] notes that:

> Increasingly it is to business that the citizen turns for fulfillment of the basic beliefs and promises of society, especially the promise of equal opportunity. Management development is little but a technical name for the means through which we carry out a central and basic part of our social beliefs and political heritage.

This view could, I suppose, be regarded as a consequence of accepting industrialization as an essential part of Western democratic process. It is certainly true that upward mobility within the industrial part of the occupational system is one means of reconciling social capacity and social position, and on this background it is easy to recognize one group of forces driving any management development scheme towards the position of becoming an institution. But what kind of institution? The problem of clarifying and reconciling conflicting "needs" and "ends" in the early stages of growth of any internal institution within the firm seems to me of immense importance. Recruitment and promotion must primarily consider the objectives and needs of the firm: attempts to rectify - by promotion - any social injustice experienced outside the firm would strike many European managers as likely to increase the total amount of injustice rather than the reverse.

The "vice" of nepotism - the unjust use of proprietorial, paternalist or personal power in recruitment and promotion - has, of course, a twin. This has no name, but it consists of insisting that performance and length of service at one level should be the main criteria for promotion to another. Let me give a discreet example, from a government service outside either of the two continents with which we are most familiar. "With us," said the head of this service, "promotion other than by seniority would be unthinkable." To accept any view of this kind - or any view which equates performance with potential - involves denial of the differing potentials of men, and it involves denial of the need to provide acceptable means of recognizing and developing those with above average potential. Earlier on, acceptance of this need was given the provocative label of the "elite" principle, although this phrase involves a special meaning of the word elite. I had better be clear that any principle of this kind does not imply that all managers should not be offered every opportunity to achieve their maximal potential. Any such principle, however, does imply that continuing equality of opportunity does not mean continuing identity of opportunity. In effect, the type of principle I am concerned to defend is related to the Victorian warning that "from him to whom much has been given, much will be expected." The problem lies in the complexity of what is "given" and of assessing this by means which can achieve the status of an institution.

Let me illustrate one of the dilemmas of management development by quoting a second American writer in this field. Discussing the institutional aspects of an enterprise, Selznick[14] says that

> . . . the formation of an institution is marked by the making of value commitments - that is, fixed assumptions, of policy makers . . . decisions affecting institutional development are critical decisions . . . and certain organizational practices . . . help to shape the key values in the organization and especially the distribution of power to effect these values . . . Decisions regarding the recruitment of personnel may become part of this critical experience . . . when selection must take account of more than technical qualifications or when leading individuals are chosen for their personal commitment to precarious aims or methods . . . where the social composition of the staff significantly affects the interplay of policy and administration, personnel selection cannot be dealt with as a routine management practice.

The views represented by the quotations from Drucker and from Selznick are widely held, and they are to me a statement of a dilemma. It may be said that the semantic issues behind this, if we took time to examine them fully, might change the definition of that dilemma. But the two views may at least give an indication of the wide range and powerful nature of one category of social forces which operate around the process of management development. The two views suggest, to go further, why

management development must either grow into an internal institution of the firm or be discredited as some new form of conspiracy to prevent "merit" achieving its "just reward." To put this point in a more ominous way, any management development scheme whose nature prevents its growth towards institutional status may easily be interpreted as no more than a new means by which the managers of a firm maintain a familiar pattern of unjust exploitation of one social group by another. On the other hand, even if a management development scheme does become an internal institution, it will, like all institutions, need careful and continuous scrutiny and adaptation, for it will inevitably be exposed to a variety of familiar risks.

The problem of reconciling the managerial needs of the firm with the social aspirations and inequalities of men may be illustrated rather more clearly if I give an almost ideal account, based on the theoretical views as well as experience of former colleagues and myself, of how a systematic scheme of management development could begin to move towards achieving an institutional character. In doing this I want to be rather detailed about points which lend themselves to consideration of certain wider issues.

Consider first the steps in introducing a scheme. Leaving aside the preceding events in the firm - they may have been tragic or comic, planned or accidental - a scheme needs ideally to be initiated as a fully integrated part of a complete program to achieve the firm's objectives. It should, that is, arise from an agreed conception of the distinctive role and competence of the firm. It needs also to be based on a fully developed conception of the personnel function. From such a conception three groups of activities will emerge: first, adapting the organization; second, manning it; and third, maintaining an effective level of morale. By morale I mean an effective attitude in the firm and its constituent groups and members towards their collective and individual tasks.

To initiate a systematic system of management development the actions needed under the heading of "adapting the organization" would be concerned with the usual forward estimates of areas of growth or contraction, based on predictions about future objectives and the future pattern of relations between the firm and the external regions earlier discussed. These estimates of future needs would involve requirements for various numbers and levels of managers, of both familiar kinds and - most important - new kinds. Planning activities to meet these estimates would come under the second area of personnel activity, that of "manning the organization." Here also would come studies of the minimal requirements in knowledge, in patterns of experience, and in levels of skill in the different groups, which make up the total management cadres, present and future. Ideally, this planning would contain what might be called "optimal flow charts" - statements of numbers and types of managers moving along

optimal career paths. Under the third component activity of the personnel function - that of "maintaining morale" - would come steps to build widespread understanding and agreement with respect to the firm's objectives, situation and managerial needs, to its available resources in potential recruits and candidates for promotion, and to its policies and practices of recruitment, development, promotion and training.

In the introduction of a scheme of systematic management development, all kinds of thorny problems would almost inevitably arise within each of these three groups of activities. Much time and patience would be needed to provide the basis for an internal institution which would first of all meet the inevitably changing needs of the firm, and, secondly, be adequately trusted and valued by a reasonably wide majority of employees at all levels. Some kind of evaluation of how far a scheme is meeting these criteria can be attempted, in a situation of managerial shortage, by a careful consideration of potential or actual managers who are attracted to the firm or who leave it. I do not mean to suggest that this type of transaction will always represent some failure. I merely suggest that in this kind of evaluation the nature and level of such transactions between the firm and the outside world is almost certain to be highly significant.

Some implicit principles and manifest problems have just been outlined in the general case. Let me now try to exemplify how a few of these have appeared in the growth of one component of a management development scheme - a central recruitment scheme for potential managers which was brought into being some fourteen years ago in its first form, and which can be described as a "boundary zone institution."

Some General Aspects of a Scheme for Recruitment to Management

The firm concerned is the United Kingdom sector of Unilever, a multi-national concern, largely concerned with domestic consumer expendibles, diversified in its field, decentralized, and with an established tradition of delegation. For these reasons the companies of which it is made up throughout the world have considerable differences of approach to such problems as recruitment to management. For immediate purposes, the main point about this recruitment component of an over-all scheme of management development is its location on what has of recent years become an increasingly important part of the boundary between the firm and its environment.

Young people are recruited to the United Kingdom units of the firm as school leavers of different kinds, from sixteen on; a little later they come in from a variety of sources. In any case, it is through promotion from levels below management that some four-fifths of the present two thousand managers in the United Kingdom sector of the concern have initially entered management. Many people would share a general feeling that, as things are, young people cannot be expected to be accurate in their

vocational guessing for some years after they have left school or university. My own experience is that if one considers the timing of the later stages of social and individual development among the intellectually more able young people in my own country, the chances of their making an effective first choice of career - or of someone else being roughly correct in agreeing with them about this - are likely to be below fifty-fifty until they are about twenty-two years of age. For reasons of this kind, what I should like to call "experimental employment" tends to go on between the ages of sixteen and, let us say, twenty-eight. In accepting this situation many firms have also come to realize that we must somehow arrange for continuing full employment, and that those now entering industry are seeking careers rather than jobs. Many firms have devised means of taking on "initial" recruits from a wide range of sources and of all ages from sixteen to twenty-eight or so. But above that age the problem of "outsiders-coming-in" begins to take on new aspects and requires a somewhat different approach. My own firm recruits in the United Kingdom both by individual companies and by central schemes, and does so both for specialists, for management training and for some levels of management. Within this over-all program, it is one central scheme of recruitment and training for general management careers that I want to describe. A more detailed definition of "general" would, of course, mean "general within a particular functional area"; or alternatively I suppose it could be defined as "non-specialist" in the sense that technical or professional qualifications play no direct or essential part in the criteria for recruitment.

The applicants ultimately selected are offered a two-year "management training" with a specific company or unit which has been arrived at by negotiation between the individual, the company chairman, and the central division concerned with recruiting. The "training" is almost entirely on-the-job, and towards the end of this period appointment to some specific management post is made. From that point on the former members of the scheme are dissolved, as it were, in the pool of young managers, and compete on even terms with them for opportunities in career development and promotion. Within this pool of young managers a differentiation between its members is made in terms of their probable age at promotion to a defined level of more senior management, and a similar differentiation is again made at this level from within the total group of senior managers eligible for still further promotion. This particular system - a chain of procedures for differentiating men of relatively high potential at three different levels in management - represents a means by which the "elite" principle is maintained. But, because any manager may join or be dropped from the "high potential" group at any one of the three levels, the system is "open" and not "closed," as would be implied by the permanent membership usually associated with the concept of an elite. Nevertheless, one point made earlier is exemplified by this scheme: it differentiates rather sharply between performance and length of service on the one hand and potential for promotion on the other.

Let me return to the initial recruitment scheme - that concerned with obtaining recruits to management of likely young people from a variety of sources. The only formal entrance qualification is that the applicant must, for this particular scheme, be between twenty and twenty-eight years old. Beyond this, they may have had any type of education, and they may be from inside the concern as well as from without. They may have been in some quite different occupation or they may have been students. At the point of inquiry or application most of them are in fact students, but that is not because academic qualifications are necessary. It is because a large proportion of our intellectually more able young people have chosen or been enabled to go on from school to university or college. For this particular scheme of recruitment the firm is at pains to stress that it is not concerned whether applicants have or have not attended technical college or university, or if so, in what faculty. More particularly, the criteria of selection boil down to the best possible guess in answer to the question: "Can we see this young man in due course as at least a middle level manager somewhere in this business?" The word "middle" is important, and there are criteria here, of course, but these criteria stem from job requirements and not from educational attainment or social background. In these elements of design you will see how some, at least, of the basic essentials for building an institution can be interpreted. To put it generally, the formal requirements for application do not derive from any past group membership or past attainment.

There are some four thousand preliminary inquiries each year. For the reasons given many of these occur at universities, often by participation in "information sessions" where exploratory discussions take place between potential applicants for the scheme, members of the central recruitment staff, and two or three young managers - for example, men who graduated from the same university three or four years ago. From these explorations, an individual interview may be asked for, and this request may be accepted or turned down by the staff concerned. All kinds of theoretical and practical questions arise about this first phase of rejection, but for our present discussion the important points are, first, that a major proportion of those who do not proceed towards application withdraw of their own volition, often because they decide against entering industry at all; and second, that the staff concerned have considerable experience with this recruitment and training process. It is accepted, nevertheless, that mistakes will inevitably occur at this stage. Some fifteen hundred applicants finally complete an application form, and from there some two hundred and fifty are invited to attend the one-and-a-half day final selection board. As a result, about eighty are offered the relevant traineeships. Some sixty to seventy accept, and the additional loss during the subsequent two-year "training" program is of the order of ten, roughly one half of whom have been invited to resign by the firm. Many kinds of questions arise from such figures. However, my purpose is not to describe or discuss this particular scheme, but to discuss the problem of fostering the

growth of a recruitment scheme towards the status of being a boundary institution. I can approach this by listing and discussing some of three criteria by which a scheme of this kind could be assessed. They are derived in part from the concept of a boundary institution. The criteria are these:

The first of these criteria - a rising or optimal share of above-average applicants attracted by such a scheme in consecutive years - is in practice surrounded by the difficult issue of "an optimal share" of the total "recruit pool" of potentially successful applicants. Nevertheless, those concerned with any recruitment scheme over years would probably accept the validity of the criterion, while recognizing the problem of expressing it in any quantified form.

I suggest this criterion of development towards institutional status, however, for the reason that selection schemes are in effect rejection schemes; they merely accept the least unlikely residue from among those who come forward. For a boundary institution of this kind the effect on those rejected is at least as important as that on those who are selected. In the scheme I am describing, managerial potential has mainly to be assessed on qualities of personality, because a relevant performance record is seldom available and in any case might well be misleading. If the "wide agreement" needed for development of a scheme with an institutional character is to be fostered, it follows that the "means" employed in these explorations of personality and potential must take into account the "ends" of the applicants as well as those of the firm. The "ends" of the applicants are multiple, but they certainly contain a desire to learn something about industrial management and about their potential for industrial careers. There is some evidence for this point in that, of the ten or so successful applicants who each year do not accept the traineeship offered to them, a majority was already considering similar offers of employment from other firms and was primarily interested in getting a "second opinion" on their potential by discovering how they would fare in a scheme whose reputation, among the external sources tapped, appears to have grown with the years. This growth in reputation has been mainly through "grapevine" or personal contacts between potential applicants and actual applicants, both successful and rejected; no "results" are published, as they are in the case of the comparable high-level government service entry which is one of the main competitors of the scheme I am describing.

There is a "seller's market" for likely potential managers - a situation likely to continue. In the scheme under discussion, only a small proportion of those first interviewed are invited to go forward to final selection boards, while two-thirds of those who reach this stage are rejected after a procedure lasting one and a half days. To meet the criterion we are discussing both the small minority of applicants accepted and the large majority rejected must regard their experiences - of initial contact,

of application, of initial interviews, and of final board - as having had at least a neutral and, if possible, an over-all positive value. In various ways the design of the procedure attempts to achieve this result. For example, each group of eight applicants to reach final boards - they are known at this stage as candidates - meets with a central staff member and a psychological adviser who will shortly conduct written psychological tests. An initial group discussion begins with a detailed description of the final components of the selection process - what will happen, exactly whom each candidate will meet, when he will be told the result, and how he can, if he wishes, inquire further about this. It goes on, as a rule, to cover jobs, careers, psychologists, psycho-analysis, tests, the results of tests, and what is or is not reported to whom. The candidates learn that they will have an opportunity to pursue these matters further in an individual interview with the psychological adviser the following day, or, as may occasionally be requested, on an entirely separate and private occasion. All this serves to diminish unreasonable tension, and the candidates settle down to the initial testing procedure while trying to see the occasion as it is: not as the beginning of a permanent judgment but as an opportunity to learn something about managerial roles and to discover how each one of them appears to four senior executives of this particular business and to two relatively neutral psychologists from outside.

On the following day, as an opening event, the eight candidates sit in a circle with the four selectors, and all twelve take part in a mutual introduction. Two selectors give a thumbnail sketch of their careers to date, the candidates follow in turn, and two selectors bring up the rear. Various aspects of this mutual introduction are of interest to the criterion of institutional development we are discussing. This is, for example, usually the first time the eight young men have met four senior executives - "captains of industry," they feel - and they are often puzzled by the difference between their expectations and what they can see of the men they meet. This is an important point of learning; the candidates see four real-life models of men in senior positions in a type of career in which they are interested, and in listening to the brief life histories of the four selectors - and in their contacts with them in later interviews - they experience a little of the personalities of these men. In consequence they can to some extent perceive how various are the people and the types of career within the firm. Conversely, they may conclude that this range is much too wide, and that they should reconsider an opening in a smaller firm. In any case, an opportunity is created for reciprocal assessment between selectors and candidates.

One further aspect of this interpersonal assessment is significant. In the mutual introduction, each candidate listens to each of the other seven candidates as they give brief accounts of themselves and their careers to date. In addition, the eight candidates will spend most of the day together in group tasks, or in privately comparing notes of their ex-

periences. Through this reciprocal appraisal of each other and on a background of the sample of senior managers provided by the four selectors, each young man can at least obtain a contribution to his emerging self-image in relation to a new world. Some degree of self-evaluation - it will not be very accurate - is at least initiated, and on this basis there can be some diminution in the feeling, almost universal in those applying for a new type of career, that the result is entirely arbitrary and mysterious. If the first of the institutional criteria we are discussing is to be met, experience of an incomprehensible outcome of selection must be kept to a minimum, and its vocational guidance aspect needs to be clearly recognizable.

The second criterion suggested as a measure of institutional growth in a selection scheme is also difficult to quantify; but in my view it is even more important than the first. The scheme operates by the use of a selectors panel of some sixty senior managers and executives, of whom four are selectors on any one occasion in the final day-and-a-half board procedure involving eight candidates. The senior men on this panel are mainly chosen on a basis of interest in the scheme, as shown by their willingness to devote time to it and to the subsequent training of those selected by it. It must also be said that some executives who are highly successful operators may occasionally be less successful in this aspect of their managerial roles. However, to a large extent the panel is representative both of the United Kingdom companies and units within which the two years training takes place and of the senior men from which it has been chosen. I should add that membership of the selectors panel will often lead to ten years' experience of selection, and this provides a basis for the second criterion of institutional status - the extent of development, in the members of such a panel, of a consensus on certain main aspects of selection. Among these are knowledge and experience of the requirements of managerial roles in relation to personality, and with this goes recognition of the nature of the skill required in the actual selection process. Selector consensus on the relation of role and personality requires agreement that managerial posts may be effectively filled in different ways by different kinds of people. The effective development of this point of view, however, requires a good deal of knowledge of the interplay of role and personality at different levels and in different units of a firm, and it involves assessment of the scope and limitations of roles as well as of individuals.

Other points of selector consensus which have a high institutional significance stem from understanding and sharing of a common basic approach to appraisal of individual personality and to early management training. This involves understanding of the significance of test and of operational intelligence levels; it involves ability to assess an applicant's probable capacity to maintain his integrity and individuality by standing by or modifying his personal perceptions in appropriate situations; and it in-

volves appraisal of the nature and extent of an applicant's capacity both to co-operate with others and to differ from them without undue offense. Within this same area of selector skill I would also place learning about the distinctive roles of selector and of psychological adviser, since two external advisers from the Tavistock Institute take part in each final board.

In the scheme described the role of the psychological advisers is initially that of clarifying the requirements of the firm, of designing the procedures, and of demonstrating to the selectors, particularly in early days, the nature of these advisory roles. Later tasks center on suggesting modification of the scheme to meet changing needs, and on offering to candidates who ask for it such counsel as may be possible on the limited knowledge which the adviser has of them. In the final board discussions, by the nature of their comments on personality, the advisers help selectors to retain an everyday vocabulary but to enlarge their frame of reference in describing what they have learned from their contact with the candidates. This also helps selectors to integrate their immediate perceptions and intuitions with their experience of trainees during the two years they spend with the companies from which the selectors come. Let me elaborate these points a little.

Towards the end of the second day of the final procedure the four selectors meet as a board with a neutral chairman - usually the head or deputy head of the central division concerned - to consider each candidate and to reach a decision about him. The psychological advisers take part in these discussions but not in the decisions. Each selector must initially give a tentative and provisional assessment of each candidate, in terms of "offer" or "no offer" or "doubtful," or occasionally in terms of potential for some appointment outside the scheme but within the firm. It is then the task of the chairman of the board to help the selectors reach a unanimous final decision. The aggregate of the provisional judgments of the four selectors will have been somewhere in the range between four "offers" and four "no offers," but even in these extreme cases there is discussion of each candidate and of the observations made about him during the procedure. From the point of view of developing, within the selectors' panel, a consensus of the kind indicated earlier, it is important for them to consider the nature of all their provisional judgments, sometimes after listening to their colleagues, sometimes by discussion of questions initiated by the president of the board, and sometimes after comments from the psychological advisers.

Equally important for our present discussion is the fact that the successful candidates are later "placed" with individual companies within the business, usually after still further thought and further interviews. Each member of the selectors' panel is likely, at any one time, to have between two and twenty of these within his particular company, and to have a considerable share in responsibility for training them towards manage-

rial appointment or for deciding that a mistake has been made. From this continuing experience selectors can be seen to develop a considerable degree of agreement on the points mentioned earlier as components of this second criterion of institutional growth in a recruitment scheme, and to the extent that a consensus on even one standard of employee begins to develop, the advantages to the firm extend well outside the recruiting scheme we are discussing. Senior managers will usually have relatively clear standards in assessing potential at the levels where they work; consensus on a standard for even a minority group at a much lower level does something to provide two end-points between which a differential scale for assessing potential can be developed.

Earlier the third criterion was defined as "reflection of standards of recruitment and the skill of the selectors panel in the degree of acceptability of the selected 'trainee' group within the firm"; and with this, of course, goes acceptance of mistakes and removal from the scheme or the firm. You will see that this criterion is largely an emergent from the first two. More specifically, it means that those selected for this training scheme shall be easily recognizable as willing learners and, at a later stage, as helpful colleagues rather than privileged competitors for promotion. The abandonment of a number of trainee schemes has largely been based on failure to meet this criterion, but in the scheme under discussion the implicit "elite" principle is carried out not only by a single early judgment at recruitment but by a linked set of appraisals at different managerial levels. Such a system of differential promotion is not "closed" but "open."

Some Final General Points

How can a boundary institution concerned with recruitment, or an internal institution concerned with promotion, be so designed and operated as to have a minimal risk of fostering narrowness of outlook and inadequacy of standards? The external selector or appraiser represents one safeguard, most familiar perhaps in the special case of the external assessor for university chairs. The analogy illustrates the need for any external individual of this kind to have an extensive knowledge of general standards. For this reason, in the case of a firm, he will usually need to come from another and parallel unit, and often from within the relevant functional area. If we follow this safeguard a little further, it would be fair to say that it ends up in the use of a panel of selectors of the kind described earlier.

In the recruitment scheme discussed one other safeguard against narrow and inadequate standards can be seen in that it is a particular task of the psychological advisers to emphasize that managerial roles can be carried out in different ways by different types of personality, and, in this connection, they must draw attention to any narrowness of judgment as

they see it. The basis of this safeguard lies in the relative width of the working frame of reference of the psychologist, for he is concerned with appraisal of personality in terms much broader than those of requirements for managerial roles. This is certainly a wider basis than that possessed by the selectors, so the psychological advisers can assist the selectors to remain broadminded about types of applicants and to be courageous about minimal levels for recruitment.

It seems likely that in recent years there has been a sharp increase in the penalty for failing to "grow" a management development scheme which can begin to satisfy the institutional criteria earlier discussed. More than ever, there is a "market in managers." The pages of advertisements for them which now appear in our ever-expanding Sunday newspapers may be regarded as no more than evidence of the rapid growth of particular industries, or of the difficulties of forward-planning in management development. There is something in this, but I prefer to compare these advertisements with the pages of analogous offers which have been obvious for half a century or more in such professional journals as those of medicine and education. To me, this "market in managers" is one aspect of the changing nature of managerial careers. It is forcing - I feel that to be the appropriate word - an accelerated change-over in the attitude of firms toward management development. Situations of heavy unemployment have been replaced by situations in which able men no longer seek jobs but careers, and a career, of course, may involve movement from one firm to another, since both firms and managers change. The experience of independent agencies concerned with these movements of managers is relevant: able managers, it appears, will seldom tolerate a firm which is ostentatiously "behind the market" in salary policy; conversely, the able manager, it is said, seldom moves for purely financial reasons. Most commonly, a positive attraction towards some new opening coincides with some loss of trust in a firm's promotion policy or with some malaise in the climate of the immediate working group. The importance of this type of self-initiated managerial turnover is that it is the more creative and the less easily satisfied managers who will tend to leave firms which are unable or unwilling to provide the adaptation and innovation needed both by them and by their managers. In firms characterized by continuous adaptive efforts and by the existence of management development schemes with some degree of institutional status it is unlikely to be the creative men who leave.

REFERENCES

1. von Bertalannfy, L. "The Theory of Open Systems in Physics and Biology," Science, Vol. II, 1950, pp. 23-29.
2. Berle, Adolph A. Power without Property. New York: Harcourt Brace, 1959.
3. Boulding, K. E. "General Systems Theory," Management Science, Vol. II. 1956, pp. 197-208.

4. Celier, C. "The Social Sciences and the Changing Role of Management," O.E.E.C. 3 Rue Andre Pascal, Paris, 1959, pp. 31-40.

5. Consumers' Association, Ltd. 333, High Holborn, London, W.C. 1.

6. Drucker, Peter. The Practice of Management. New York: Harper and Brothers, 1954 and London: Heinemann, 1955.

7. Eells, R. The Meaning of Modern Business. New York: Columbia University Press, 1960.

8. Foote, Nelson. "The Autonomy of the Consumer," Consumer Behavior. New York: New York University Press, 1954.

9. Harbison, Frederick and Charles A. Myers. Management in the Industrial World: An International Analysis. New York: McGraw-Hill Book Company, Inc., 1950.

10. Hurff, G. B. Social Aspects of Enterprise. Philadelphia: University of Pennsylvania Press, 1950.

11. Katona, G. The Powerful Consumer. New York: McGraw-Hill Book Company, Inc., 1960.

12. Levitt, Theodore. "Marketing Myopia," Harvard Business Review. July-August, 1960, pp. 45-56.

13. McGregor, Douglas. The Human Side of Enterprise. New York: McGraw-Hill Book Company, Inc., 1960.

14. Selznick, P. Leadership in Administration. Evanston, Illinois: Row, Peterson and Company, 1957.

15. Trist, Eric L. and F. E. Emery. "Socio-Technical Systems," Management Sciences, Models and Techniques, Vol. II. London: Pergamon Press, Ltd., 1960.

Victor L. Andrews[1]
Graduate School of Business Administration,
Harvard University

Captive Finance Companies:

Their Growth and Some Speculations on Their Significance

I. INTRODUCTION

Recent sharp growth in numbers and assets of captive finance companies[2] is of economic interest on two principal scores: (1) it is the debut of a group of large financial intermediaries, and (2) it has produced assertions that the captives represent an advance in corporate financial technique. Several topics within these broad areas are significant. Under (1) we might point out that, like the kindred commercial and sales finance companies, the captives intermediate between the capital and money markets and nonfinancial borrowers. The captives are differentiated from the independent finance companies in that many lend on security tolerated at best and often rejected by the independents, their lending buttresses the sales goals and credit policies of the parent companies, and they probably assume greater exposure to risk. The assertion under (2) rests upon an "advantage" which allegedly stems from the ability of a captive to finance a portfolio of notes receivable with heavier proportions of debt than could have been used if funds had been drawn from a conventional capital structure of a nonfinancial parent corporation.

This paper is a progress report from continuing research on captive finance companies. Emphasis is placed here upon the first of the two

---

[1] Assistant Professor of Business Administration, Harvard University. Statistics presented in this article were assembled while I was a member of the faculty of the School of Industrial Management at MIT. Research was generously financed by the Sloan Foundation. A succession of research assistants placed me in their debt, and although numbers prevent individual recognition, it is a pleasure to acknowledge their help. A number of colleagues offered helpful criticisms of an earlier draft, and I am grateful for their contributions.

[2] For purposes of this paper captive finance companies (a) are defined to be financial subsidiaries of nonfinancial parent corporations dealing in greater than local markets with (b) assets composed predominantly of notes receivable. The former requirement eliminated the many local captives of automobile dealers and merchants, and the latter eliminated other subsidiaries which are primarily financial but unconnected with the kind of lending considered here.

broad areas of interest outlined above; that is, captive finance companies as suppliers of funds. Sections II and III present, respectively, a brief background of the captives' operations and quantitative data. Speculation in the last two sections attempts to show the need for exploration of the captives' characteristics as lenders and the need to test their alleged superiority as devices for financing credit extension by nonfinancial corporations. These last two sections are both an invitation to more research and a chart of its early steps.

To the best of my knowledge this is the first time that quantitative data for a large sample of these companies have been gathered and published. There has, however, been considerable notice taken qualitatively of the growth and function of the captives.[3]

II. OPERATIONS OF THE CAPTIVE COMPANIES

Marketing Policy Pressure

The captive companies finance the inventories of distributors and/or dealers and/or they hold the notes receivable generated by final sales. Their assets are, therefore, almost entirely wholesale financing, i.e., claims secured by inventory, or retail installment receivables.

A common feature of the subsidiaries is their involvement with the marketing of goods which are high priced relative to the resources of the merchandiser or ultimate owner, and sometimes both. In producers goods, for illustration, we see captives financing inventories and/or sales contracts of truck trailers, light aircraft, commercial air conditioning, telephone switchboards, farm machinery, earth moving equipment, and commercial laundry machinery. In consumer goods, captives finance automobiles, mobile homes, consumer appliances, prefab housing, tires, home furnishings, music instruments, and the great variety of goods marketed by the department store - mail-order houses.

The subsidiaries vary greatly in their purposes and lending, but we may profitably oversimplify and describe those of manufacturers and those of retailers as archetypes.

Manufacturers' Subsidiaries

Little imagination is required to envisage how financial pressure is transmitted backward through the marketing apparatus from the final purchaser and/or dealer to the manufacturer in product lines such as those named above. Thinly financed buyers, whether businesses or consumers, are susceptible to the blandishments of installment credit. This invites

[3] Ref. (12), pp. 489-490, 495, (2) and (11).

competitive credit extension and resultant burdensome receivables portfolios, which the distributive trades often cannot shoulder.

To the trade industry's problem of financing receivables is added the necessity of carrying high unit value inventories, a demand for funds often severely accentuated by seasonal peaks.

The range of sources of funds on which trade businesses may draw is typically not great. Access to the financial markets for this group frequently consists of bank borrowing and little else. Unincorporated businesses, which are a high proportion of the trade industry, cannot draw upon the securities markets, save perhaps through commercial and residential mortgages, and access for small corporations is relatively expensive when it exists. The market for direct placements and term loans is the most likely source of long-term funds for small corporate borrowers, but even through this method the obstacles to borrowing on the small scale of most local distributive corporations are great. In short, the problems of long-term finance for small business prevail.

I offer no proof now, but circumstances of the 1950's could have converged to channel pressure from the trade industry to manufacturers. Either or both the thrust of the inventory accelerator and receivables growth during a period of sales expansion could have forced trade businesses to seek external financing. A bottleneck in access to long-term funds would have shunted demand to short-term lenders. If other short-term lenders were unable or unwilling to provide funds to the trade industry, the finger could have pointed at manufacturing suppliers.[4]

In any event, the manufacturer's captive finances its parent's products through the marketing chain. Inventory in the hands of distributors and dealers is secured to the subsidiary by one of the conventional title retention devices. Some few manufacturer's subsidiaries extend only inventory financing, but more commonly they also discount the notes receivable generated by sale to the final purchaser. Discounting is done under a variety of arrangements, some with recourse, some without. Usually these receivables are generated by sale of the parent company's products, but fear of action under the Robinson-Patman Act has compelled the larger captives in consumer appliances to accept any receivable. Some examples are known of active expansion by captives into product lines of other companies.

---

[4]Some research is cited below which argues that manufacturing corporations have indeed functioned this way in recent financial episodes.

Without pretense at universality of application we might quote from an internal company publication to illustrate.

> Generally the terms of "credit type" sales of new and used trailers by the (confidential) . . . to its customers, either direct or through dealers . . . vary in the amount of down payment required . . . The installment notes represent the balances due on these sales after these down payments, and are secured by recorded conditional sales contracts . . .
>
> (The financial subsidiary) has full or partial recourse against dealers for balances owing in sales financed by a dealer through the company.
>
> The floor plan agreement provides that, if the equipment is sold during the period covered by the note, the unpaid balance shall become due and payable. The notes are secured by recorded trust receipts and unrecorded chattel mortgages or conditional sales contracts . . .

Trade Industry Subsidiaries

A financial demand posed by a different kind of marketing problem has led to formation of captives by trade corporations. The department store - mail-order houses which extend installment credit on consumer hard and soft goods have been faced with burgeoning portfolios of receivables produced by rising sales, by liberalization of credit terms and by sales composition shifting in the direction of installment purchases. Thus, the retail trade captive is an outgrowth of sales policy of the parent just as is the subsidiary of a manufacturer, but in retailing the captive need not finance inventory through intermediate stages of distribution. Its primary asset, therefore, is a consumer installment note receivable.

A quotation from a prospectus of the Sears Roebuck Acceptance Corporation (SRAC) will illustrate.

> The Company (SRAC) has entered into an Agreement with Sears (the parent) . . . under which Sears sell and assigns to the Company all conditional sales contracts . . . arising out of merchandise sales made by Sears in such retail and mail-order outlets as are mutually agreed upon from time to time.
>
> All such sales of Contracts to the Company are without recourse against Sears . . .

III. GROWTH OF COMPANIES AND ASSETS

Number of Active Subsidiaries

The author has identified 102 financial subsidiaries of 90 parent corporations active in the post-World War II period through 1960.[5] Table 1 is a breakdown by two-digit industry of the number formed in each year. The concentration of subsidiaries in the high-ticket, hard goods industries is evident. Nonelectrical machinery, including the earth moving and agricultural equipment makers, had the greatest number at the end of 1960 with 20 captives. Transportation equipment with the auto, truck, and truck trailer producers included had the second largest number with 18. The appliance makers in electrical machinery were primarily responsible for the 16 captives in electrical machinery. In total these three industries, then, accounted for more than half the active subsidiaries formed after World War II. Lesser concentrations exist with the cotton-seed producers in food and kindred products and in retail trade.

A second conspicuous characteristic of these captive company births is their heavy concentration in the years 1953 to 1957 and to a reduced extent afterward through 1960. At the outset of the postwar era 14 subsidiaries were active. Relatively few additions were made through 1952. To the total of 29 existing then, however, 56 were added through 1957 to raise the figure to 85. Twenty-one more captives were added between 1957 and 1960, which was a pace sharply reduced but still materially higher than that of the early postwar years.

Growth of Companies and Assets

From various sources complete balance sheet data[6] have been assembled and aggregated for 65 companies operating between the end of World War II and the end of 1960. Table 2 shows these data in dollars and a percentage breakdown of total assets and liabilities. The extreme right-hand column of the table is a record of the number of active companies tabulated in each year. The number in any year differs from the total because of deactivations and some few discontinuities in availability of figures for small companies.

---

[5] A statement of the means used and an elaboration of criteria employed for inclusion appear in Appendix A.

[6] "Complete balance sheet data" are defined to be balance sheets for every year of a company's existence after World War II through 1960. For example, General Electric Credit Corporation has existed since the 1930's so that complete data would be comprised of balance sheets for every year in the period 1946-1960 inclusive. On the other hand, for a company which had been formed, say, in 1955, complete data would be balance sheets for the period 1955-1960. Companies which discontinued activity before 1960, of course, would be included for only those years of their active life. For a description of sources and methods of collection see Appendix B.

TABLE 1

POSTWAR FORMATION DATES OF NATIONALLY SIGNIFICANT CAPTIVE FINANCE COMPANIES BY TWO-DIGIT INDUSTRY CLASSIFICATION

SIC Two-Digit Industry	Pre-1946	1946	1947	1948	1949	1950	1951	1952	1953	1954	1955	1956	1957	1958	1959	1960	Total
20 Food and kindred products	2					2			2					1			7
22 Textile mill products										2		1					3
24 Lumber and wood products			1						1	1	1	1					7[1]
25 Furniture and fixtures	1						1				1						3
29 Petroleum refining and related industries	1										1						2
30 Rubber and miscellaneous plastic products	1	1								1							3
31 Leather and leather products					1	1											2
32 Stone, clay and glass products	1																1
34 Fabricated metal products				1							1		2				4
35 Machinery, except electrical	1				1	1				2	2	3	5	2	3		20
36 Electrical machinery	2						2	1	2	3	1	3	1		1		16
37 Transportation equipment	4			1					1	2	3	3		1	1	2	18
38 Instruments				1					1	1	1						4
39 Miscellaneous manufacturing												1	1			1	3
50 Wholesale trade										1	1					1	3
53 Retail trade	1											2	1			1	5
70 Hotels, etc.														1			1
Total	14	1	1	3	2	4	3	1	7	13	12	14	10	5	5	5	102[1]
Cumulative total	14	15	16	19	21	25	28	29	36	49	61	75	85	90	95	100	102[1]

[1]Two companies formed at unknown dates.

Total assets, shown in column 6, rose from the presumably war-shrunken base of $0.3 billion in 1946 to $9.2 billion in 1960. Since GMAC is a case apart, column 7 shows total assets less those of GMAC, and is in one sense more representative. Because the absolute base of 1946 was small, proportionate expansion of non-GMAC assets through 1953 was sizable, but in absolute terms had reached only $0.8 billion. Thereafter, absolute expansion was impressive. A net addition of $3.1 billion was made to non-GMAC assets through 1960 to raise the total to slightly under $4.0 billion. Paralleling the growth of the number of companies, much of this came during the 1955 to 1957 boom and in 1959 and 1960.

The $6.9 billion of retail notes receivable outstanding[7] in 1960 equaled almost three-fourths of the total. In the period since 1952 this proportion ranged in the 70 per cents. GMAC held $3.9 billion of the retail financing outstanding in 1960. Wholesale or inventory financing mounted to $1.8 billion in 1960 from almost a negligible absolute level at the outset of the postwar period.

From the foregoing it is clear that other kinds of assets made up only minor portions of the total. Except for immediate postwar years, cash and cash equivalents were low fractions of total assets. Although balance sheet condensation in Table 2 eradicates the distinction between cash and securities holdings, the author's worksheets show that securities were virtually negligible - less than 0.4 per cent at the end of 1960. Miscellaneous assets equaled 1.6 per cent of the total in 1960, and fluctuated below this figure in all the years shown here. This category is composed of term loans, leased assets, and a wide assortment of other assets, all negligible in total, though any one may be important to an individual company.

The captives draw the bulk of their financing from the debt market. Over the postwar years reliance on various types of debt shifted noticeably. In the years through 1952 short-term borrowing, consisting of bank loans and commercial paper, comprised the single most important category of debt financing. Except for 1946 total short-term debt ranged in this period between 46 per cent and 53 per cent of total sources. Thereafter, it comprised between three-tenths and four-tenths of total financing.

Commercial paper tended to be the largest of the two components of short-term borrowing. Through 1953 it was not uncommon for com-

---

[7] The term "retail financing" is not synonymous with consumer credit. Final sales of procedure goods are classified here as retail financing. This would be called equipment financing or time sales financing in many discussions. In fact, much of retail financing by the captives is consumer credit in the electrical machinery industry, all of it in retail trade, and most of that done by GMAC in transportation equipment is consumer finance.

TABLE 2

ASSETS AND LIABILITIES OF A SAMPLE OF CAPTIVE FINANCE COMPANIES, 1946-1960
(Dollars in Millions)

ASSETS							
1	2	3	4	5	6	7	8
Years	Cash and Securities	Wholesale Notes Receivable	Retail Notes Receivable	Miscellaneous Assets[1]	Total Assets[2]	Total Assets Less GMAC	No. of Cos. Included
1960	$417.7	$1796.7	$6857.4	$154.5	$9226.3	$3961.7	56
1959	377.4	1376.6	5482.4	175.0	7411.5	3143.5	54
1958	323.0	1342.9	4581.5	147.7	6395.0	2460.4	54
1957	310.3	1354.2	4678.7	149.5	6492.7	2048.3	49
1956	239.5	1038.3	4228.2	122.9	5629.0	1555.3	41
1955	206.1	936.7	3707.1	102.9	4952.9	1117.3	34
1954	167.2	621.1	2673.8	65.1	3527.3	851.4	26
1953	155.9	540.0	2506.6	53.3	3255.9	781.8	20
1952	119.7	304.5	1799.7	32.2	2256.1	521.4	17
1951	143.4	454.5	1398.6	34.3	2030.9	578.4	16
1950	138.0	439.2	1405.7	29.4	2012.3	464.5	15
1949	97.2	252.4	1025.9	20.4	1396.1	319.6	15
1948	77.4	155.9	595.5	12.8	805.5	204.2	13
1947	64.9	120.7	303.4	8.2	497.2	106.0	10
1946	63.8	83.3	137.4	6.9	291.4	46.6	10
			(PER CENT OF TOTAL ASSETS)				
1960	4.5	19.5	74.3	1.6	100.0		
1959	5.1	18.6	74.0	2.4	100.0		
1958	5.1	21.0	71.6	2.3	100.0		
1957	4.8	20.8	72.0	2.3	100.0		
1956	4.3	18.4	75.1	2.2	100.0		
1955	4.2	18.9	74.8	2.1	100.0		
1954	4.7	17.7	75.8	1.8	100.0		
1953	4.8	16.5	76.9	1.6	100.0		
1952	5.3	13.5	79.7	1.4	100.0		
1951	7.1	22.4	68.9	1.7	100.0		
1950	6.8	21.8	69.8	1.5	100.0		
1949	7.0	18.1	73.4	1.5	100.0		
1948	9.6	19.3	69.4	1.6	100.0		
1947	13.1	24.3	61.0	1.6	100.0		
1946	21.9	28.5	47.1	2.3	100.0		

[1] Includes term loans, leases, mortgages and a great miscellany of other assets.

[2] Detail will not add to total due to rounding.

TABLE 2 (continued)

LIABILITIES										
1	2	3	4	5	6	7	8	9	10	11
	Short-Term Borrowing					Long-Term Debt				
Years	Bank Loans[1]	Commercial Paper	Total	Accruals	Borrowing From Parent Cos.	Unsubordinated	Subordinated	Total	Loss Reserves	Equity
1960	1992.2	1912.4	3904.6	330.2	429.4	2842.6	676.4	3518.9	100.8	949.4
1959	1261.0	1484.7	2746.6	302.3	456.1	2602.0	536.3	3138.2	82.7	682.8
1958	927.0	1085.0	2011.9	278.2	501.3	2467.8	475.5	2943.2	71.9	588.4
1957	1024.0	1327.0	2351.0	283.3	400.8	2471.8	433.0	2904.8	71.2	482.7
1956	772.4	1116.4	1888.8	276.1	393.5	2199.5	428.0	2627.5	60.0	382.7
1955	802.3	928.9	1731.2	252.0	300.4	1890.4	381.4	2271.8	49.6	347.7
1954	469.7	756.7	1226.4	203.1	276.8	1221.5	288.2	1509.7	37.9	273.3
1953	478.4	834.3	1312.7	180.3	154.4	1086.4	250.0	1336.4	30.5	241.4
1952	446.4	749.4	1195.8	130.1	146.7	442.1	105.0	547.1	22.6	214.8
1951	489.0	504.4	993.4	112.5	164.8	436.5	105.0	541.5	18.7	199.9
1950	588.8	337.2	925.0	97.9	266.9	406.2	105.0	511.2	16.1	194.2
1949	419.8	310.4	730.3	61.9	159.1	208.7	80.0	289.7	13.3	142.8
1948	146.3	230.8	377.0	31.2	94.1	173.1	3.0	176.1	9.2	117.8
1947	102.2	125.0	227.2	15.5	127.9	9.0	-	9.0	7.5	110.1
1946	33.5	38.6	72.1	7.9	91.2	10.0	-	10.0	6.4	103.9
					(PER CENT OF TOTAL LIABILITIES)					
1960	21.6	20.7	42.3	3.6	4.6	31.8	7.3	38.1	0.9	10.3
1959	17.0	20.1	37.1	4.1	6.2	35.1	7.2	42.4	1.1	9.2
1958	14.5	17.0	31.5	4.8	7.8	38.6	7.4	46.0	1.1	9.2
1957	15.2	20.4	36.2	4.4	6.2	38.1	6.7	44.7	1.1	7.4
1956	13.7	19.8	33.6	4.9	7.0	39.1	7.6	46.7	1.1	6.8
1955	16.2	18.8	34.6	5.1	6.1	38.2	7.7	45.9	1.0	7.0
1954	13.3	21.5	34.8	5.8	7.9	34.6	8.2	42.8	1.1	7.7
1953	14.7	25.6	40.3	5.5	4.7	33.4	7.7	41.0	1.0	7.4
1952	19.8	33.2	53.0	5.8	6.4	19.6	4.7	24.3	1.0	9.5
1951	24.1	24.8	48.9	5.5	8.1	21.5	5.2	26.6	1.0	9.9
1950	29.2	16.8	46.0	4.9	8.3	20.2	5.2	25.4	0.8	9.7
1949	30.1	22.2	52.3	4.4	11.4	15.0	5.7	20.7	0.9	10.2
1948	18.1	28.7	46.8	3.9	11.7	21.5	0.3	21.9	1.1	14.6
1947	20.5	25.2	45.7	3.2	25.8	1.8	-	1.8	1.4	22.1
1946	11.3	13.4	24.7	2.8	31.8	31.3	3.4	-	2.0	35.7

[1] Includes a negligible amount of borrowing from finance companies by one captive in the years 1957-1960.

mercial paper to comprise more than one-fourth of total financing, but in the years thereafter the fraction was close to one-fifth. No breakdown of the total into directly placed and publicly sold paper is available yet, but the author has the impression that most is sold directly.

The tendency for short-term debt to decline in proportion to total financing also shows in the proportion of bank loans in total sources. Before 1953 bank loans ranged between 20 per cent and 30 per cent of all sources of funds, but afterwards fluctuated between 15 per cent and 20 per cent. Whereas long-term debt comprised in the neighborhood of one-fourth or less of total financing through 1952, it remained above two-fifths of the total in the years 1953 to 1959. The bulk of long-term indebtedness is and has been unsubordinated.

Except for the early post-World War II years borrowing from parent companies supplied a relatively small amount of total financing. A minor amount is subordinated. It is not proportionately significant in the aggregate, though for individual companies it sometimes plays an important role.

Equity committed to the captives in this sample - by the parent companies alone save for one exception - equaled more than 10 per cent of total sources in 1960. Loss reserves comprised in the neighborhood of 1 per cent of total sources.

Industry Distinctions

Breakdowns of captive company assets in four industries appear in Table 3. The industries covered are nonelectrical machinery, electrical machinery, transportation equipment, from which GMAC has been subtracted, and retail trade. These four industries accounted in 1960 for $3.6 billion of the $3.9 billion of non-GMAC assets.

Assets held by captives in nonelectrical machinery climbed steeply from a nominal level in 1952 to $1 billion in 1960. The recessions of 1954 and 1958 interrupted or slowed the rise. Wholesale notes receivable ranged mostly between 50 per cent and the low 60 per cents. In absolute terms wholesale financing rose with cyclically booming sales in 1950 to 1951, 1953, and 1955 to 1957, presumably because distributors were building inventories under the thrust of the inventory accelerator. Retail notes receivable were, of course, the other major asset, ranging in the years of sustained growth between one-fourth and over two-fifths of assets. Since 1955, there has been a strong tendency for retail financing to expand relative to the total.

In electrical machinery, with the numerous captives of the consumer appliance manufacturers, the advance of total assets was more even

and dated farther back than in the case of nonelectrical machinery. The second half of the postwar period, however, witnessed greater growth as assets rose from slightly under $0.3 billion in 1953 to $1.2 billion at the end of 1960. Heavy consumer sales by the appliance makers resulted in a preponderance of retail notes in total assets throughout the period under study, but in the years since 1955 wholesale notes rose in relation to total assets from 7.4 per cent in 1954 to 15 per cent in 1960.

The captives in transportation equipment (sans GMAC) had a rising level of total assets extending back to the immediate postwar period, but cyclical influences twice interrupted year-to-year growth. From a negligible $17 million in 1946, assets climbed to over $0.8 billion in 1960. In the latter half of the postwar era retail notes hovered in the vicinity of four-fifths of total assets and wholesale notes were roughly steady around 10 per cent.

Captives for which we have data in the retail trade are relative latecomers. Since 1957, assets, composed almost entirely of retail notes receivable, rose from $153 million to $621.7 million.

Receivables Held by Captives Related to Reported Statistics

With some very few exceptions the Quarterly Financial Report for Manufacturing Corporations reports receivables held only by nonfinancial parent corporations. It is of interest, therefore, to see the receivables held by credit captives in manufacturing in relation to reported figures in Table 4.

The captives' receivables climbed relative to reported receivables in all three manufacturing industries shown over the postwar period. Captives in nonelectrical machinery held a sharply ascending fraction after 1956, but in electrical machinery and transportation equipment growing significance of the captives' holdings dated back to 1953 or 1954. By the end of 1959 receivables held by captives in nonelectrical and electrical machinery had risen to over one-fifth of those held by nonfinancial corporations. By the end of 1960 the figure was almost 19 per cent with GMAC removed from the captives in transportation equipment.

There seems to be good ground for thinking that research results predicated on the receivables reported in the Quarterly Report have suffered from understatement. This comment applies with particular force to recent research which has depended upon receivables holdings of corporate manufacturers as a measure of gross credit flows to other segments of the economy.

TABLE 3

DISTRIBUTION OF ASSETS OF A SAMPLE OF CAPTIVE FINANCE COMPANIES IN FOUR INDUSTRY GROUPS, 1946-1960

(Dollars in Thousands)

SIC Ind. No.	Name	Year	Cash and Securities	Wholesale Notes Receivable	Retail Notes Receivable	Miscellaneous Assets	Total Assets
35	Machinery, except electrical	1960	$74,144	$410,350	$507,593	$ 7,472	$ 999,569
		1959	68,955	480,532	402,282	5,616	957,385
		1958	59,900	380,540	284,379	5,455	730,274
		1957	47,776	321,275	150,552	3,321	522,924
		1956	36,628	208,442	91,405	1,747	338,222
		1955	29,395	155,853	63,613	5,780	254,641
		1954	31,410	118,596	51,263	3,572	204,841
		1953	31,447	136,191	40,350	5,433	213,421
		1952	9,900	-	26,077	354	36,331[1]
		1951	22,163	111,903	22,531	7,243	163,840
		1950	20,292	67,207	24,757	3,180	115,436
		1949	7,466	35,917	19,936	1,806	65,125
		1948	10	-	1,689	100	1,799
		1947	10	-	1,687	100	1,797
		1946	10	-	1,668	100	1,778
36	Electrical machinery	1960	89,101	178,849	894,719	19,651	1,182,320
		1959	69,097	218,724	633,585	14,348	936,236
		1958	62,731	134,609	470,883	10,194	678,417
		1957	47,229	102,023	442,183	4,574	596,189
		1956	32,268	67,144	417,630	4,423	521,465
		1955	26,434	46,298	311,729	3,306	387,767
		1954	24,417	22,674	260,020	1,740	208,851
		1953	24,573	14,835	249,697	2,290	291,487
		1952	20,052	10,811	180,894	1,481	243,238
		1951	10,467	27,661	117,460	1,055	156,643
		1950	11,617	8,632	130,087	784	151,120
		1949	11,091	6,816	103,498	684	122,057
		1948	10,154	6,336	75,248	209	91,957
		1947	4,795	3,789	45,449	107	54,146
		1946	2,199	1,814	18,652	55	22,730

[1] Drastic shrinkage of assets in this year is attributable to a one-year virtual deactivation of one large captive, possibly because of Korean war production by the parent.

TABLE 3 (continued)

SIC Ind. No.	Name	Year	Cash and Securities	Wholesale Notes Receivable	Retail Notes Receivable	Miscellaneous Assets	Total Assets
37	Transportation eqpt. minus GMAC	1960	$47,366	$114,803	$648,959	$26,276	$ 837,404
		1959	35,425	62,922	487,340	18,873	604,560
		1958	37,986	55,798	472,141	25,111	591,036
		1957	42,647	56,722	506,139	28,617	634,125
		1956	39,094	68,178	469,534	19,780	596,586
		1955	16,242	60,489	320,181	6,680	403,592
		1954	13,410	26,450	235,263	4,899	280,022
		1953	11,013	28,793	204,475	3,242	247,523
		1952	15,668	34,728	196,230	2,496	249,122
		1951	22,821	40,511	173,115	4,456	240,903
		1950	15,553	28,514	134,190	2,864	181,121
		1949	12,368	18,826	87,860	2,015	121,069
		1948	11,138	17,945	70,262	975	100,320
		1947	3,516	7,967	28,729	1,120	41,332
		1946	3,475	3,095	9,468	1,289	17,327
53	Retail trade	1960	19,590	-	597,870	4,267	621,721
		1959	2,999	-	339,653	25,908	368,560
		1958	2,800	-	262,932	2,735	268,467
		1957	3,700	-	145,609	4,207	153,516

TABLE 4

RECEIVABLES HELD BY CAPTIVES[1] AS PERCENTAGES OF RECEIVABLES REPORTED FOR NONFINANCIAL CORPORATIONS IN THREE INDUSTRIES IN QUARTERLY FINANCIAL REPORT FOR MANUFACTURING CORPORATIONS AT DECEMBER 31

Industry	1960	1959	1958	1957	1956	1955	1954	1953	1952	1951	1950	1949	1948	1947
Machinery, except electrical	22.0%	21.5%	17.4%	11.3%	7.3%	6.8%	6.7%	6.5%	0.8%	5.6%	4.9%	4.1%	-	-
Electrical machinery	25.7	20.8	18.1	18.8	17.2	16.0	14.4	13.9	10.0	11.2	12.4	14.9	10.0%	6.4%
Transportation eqpt. sans GMAC	18.8	13.8	14.7	15.8	15.5	12.2	9.4	8.4	8.4	9.5	10.3	10.3	7.8	3.6

[1] As shown in Table 3.

IV. THE CAPTIVES AS A PART OF BROADER FINANCIAL DEVELOPMENT

Growth in number and total assets of the captives is interleaved with broader developments in the financial markets. The discussion here relates the timing of the captives' burgeoning to the behavior of inter-business financing during recent business fluctuations and relates the character of their loans to maturation of the financial system. Emphasis is placed on the latter, but evidential limitations make the argument mainly heuristic.

The Captives' Loans as a Specie of Trade Credit in Recent Cycles

Trade credit has been recognized as an important element of inter-firm financing and a responsive, flexible credit flow far back in our history, and evolution of a number of devices for short-term lending traces back to its usage. Sayers and Foulke noted the role played by mercantile credit during the 19th century.[8] Lavington noted the function of trade credit in Great Britain in the early 20th century,[9] and for the American institutional environment Jacoby and Saulnier[10] and also Goldsmith[11] have considered the interaction of bank lending and trade credit in financing business for earlier periods.

The particular conditions of the fifties evoked their own response. Research has suggested that the flow of net trade credit[12] from the corporate universe during the 1955 to 1957 boom financed smaller and/or unincorporated businesses,[13] and was spurred by shortage of supply in the bank loan market. It is impossible to clinch the point that the credit flowed systematically from the larger corporation universe to smaller businesses, but balance sheets suggest this.[14] During the 1955 to 1957 boom, corporate manufacturers reduced their liquidity in association with rising short-term

---

[8] Cited in Ref. (9), footnote 2.

[9] Ref. (8), Chapter XLII.

[10] Ref. (7), p. 960.

[11] Ref. (6), p. 215 and 247.

[12] Meaning that trade receivables exceed trade payables, implying that loans extended via receivables exceed loans received via payables.

[13] Ref. (4), pp. 4-5, and Ref. (12), pp. 25-26.

[14] Data from the Statistics of Income by size of firm do not break down notes and accounts payable. Thus, it is impossible to net trade payables directly against receivables for breakdowns by size of firm.

interest rates and tightening bank reserve positions.[15] Also, the ratio of net receivables to sales rose in association with growing sales for corporate manufacturers with assets over $50 million in the 1955 to 1957 period. Coupling these observations, it can be inferred that sales support was derived from liberalization of credit terms amounting, presumably, to extension of credit to mostly nonmanufacturing businesses.

A continued rise of the net receivables/sales ratio for manufacturing corporations after the years of peak interest rates and during periods of monetary ease vitiates the plausibility of the contention that rationing in the bank loan market was responsible for the growth of net receivables outstanding.[16] Correlation of large manufacturers' sales growth with lengthening net receivables may exist for the years after 1957 as it did for 1955 to 1957. If so, competition among large firms at the level of credit extension may bear the greater burden of explanation. A firm with a comfortable liquidity position at the outset of a period of credit shortage will be affected primarily through restraint upon the ability of prospective customers to make purchases, rather than by direct impact of policy restraint. Of course, limited lending capacity of the banking system and reduced liquidity of other business firms would have provided the fertile ground upon which these policies fell..

The growth of the credit captives may have been a response to the same forces leading to these developments. Peak rates of captive company formation and asset growth coincide in time with expanding gross credit flows from corporations. Size and industry of parent company lenders and borrowers through traditional trade credit and through captives correspond closely. Industry clustering of the captives may be explicable in that credit terms loosening throughout an industry would presage growing receivables stocks and force adoption of an expedient in financing technique, viz., a credit captive.[17]

Confirmation of the linkage of captive companies with the forces which have produced growth of net receivables more generally in the last half-dozen years must await probative testing at the level of the firm.

---

[15] Ref. (9), pp. 430-431.

[16] Ref. (5), pp. 3-4.

[17] Though this is anything but clearly true. If a captive is conceived of only in its role as a financing device, its advantages would presumably apply to receivables portfolios of any length. Of course, a portfolio marked by slowing turnover would accentuate the problem of the financial manager, and perhaps provoke a response after some threshold level had been reached. Thus, the captives would be associated in individual firms with receivables growing relative to sales and probably relative to total assets.

The Captives as a Refinement of the Capital Market

A capital market functioning perfectly in a world with risks and lending costs would produce interest rate differentials traceable to their influence. Loans of equivalent risk would bear the same rate of interest throughout the economy as the mark of the market's allocational efficiency. Competitive pressure would assure the market's operating efficiency. Rationing and nonactuarial yield differentials between equivalent risk and equal cost loans would be symptomatic of imperfection of the market.

Broadly, then, the captives improve upon the mechanism of funds allocation if they remove or reduce uneconomic rate differentials, alleviate rationing, or cut borrowing costs by improving operating efficiency. Visible evidence of market refinement would be arbitrage by the captives between the capital and money markets on one hand and borrowers on the other hand who are unable to draw directly upon the same markets or unable to do so at rates better than those offered by the captives. In this role the captives would function as the independent finance companies and leasing companies do. Reduction of loan costs will not be visible externally.

Some Possible Sources of Novelty in the Captives' Lending

Some of the captives' capacities as risk evaluators and bearers suggest that they are geuninely innovative lenders, and that their ability, through joint use of facilities and personnel, to exploit semi-fixed and fixed costs of the parent makes it possible to achieve cost reductions in loan administration. It is argued sometimes that the perspective of local bankers and trade suppliers and their access to local information enables them to evaluate more accurately and accept risks which are exaggerated in the eyes of more remotely situated lenders. The manufacturers' captives share this advantage, and add others through their association with distributors and dealers handling the parent company's line of goods. Usually, field representatives of a parent manufacturer do double duty in their visits to dealers and distributors in checking the physical security underlying inventory lending. Also, as a result of the sales connection between parent manufacturer and marketing apparatus, the captive is automatically privy to a certain amount of information about the financial condition of dealers and distributors ordinarily obtainable by outside lenders only at the significant cost of repeated audits and sustained surveillance of the quality of receivables.

Since a captive contributes to sustaining its parent company's sales rate, a part of its profit is buried in the parent's profit figure. It follows that a captive could manipulate interest cost and other terms of its lending for the sake of increased manufacturing or merchandising profit. Thus, increased assumption of risk and its fruits may be difficult to trace.

Another nuance is the favorable situation of the captives in some manufacturing industries regarding their ability to repossess without loss or with minimal loss. Banks are unwilling to lend on heavy equipment when the second hand market is thin, and this is often the case in sparsely populated areas. Since the parent can easily recondition and resell repossessed equipment, however, the captive is ideally postured to assume this risk. Risk is minimized for the captive lender on retail paper too. When default occurs on retail paper, banks and independent sales finance companies have little recourse except repossession with all the hazards it entails but, through the compulsion of franchises and, perhaps, lack of alternatives open to dealers in financing, the captives often are able to exact partial or complete recovery from the originating dealer.

Fragmentary Inferences About the Character of the Captives Lending

Solely on the basis of the types of loans made by the captives there is no ground for asserting that they are innovative lenders. The extent of the captives' lending on secured forms, however, argues otherwise. The development of secured lending has been crucial in widening the availability of funds to credit risks otherwise unacceptable to lenders. However, inventory has played a relatively minor role. In fact, it is treated usually as an unwanted stepchild, a lead to profitable receivables financing. Equipment financing is more marketable, but by comparison with previously existing sources, the captives have emerged as quantitatively significant lenders on this type of security as well as inventory. For illustration I will rely solely upon the considerable data accumulated by the survey of credit and capital sources conducted by the Federal Reserve System early in 1958 and upon surveys of member bank loans to business.

An interview survey of representatives of more than 670 commercial banks indicated in 1958 that approximately one-half made loans secured by inventory "relatively frequently." Significantly, perhaps, small banks in small communities reported less inclination to make relatively frequent use of inventory loans than their larger brethren in larger communities.[18]

The last date for which we have a cross-section of secured member bank loans outstanding is October 1955. Of $30.8 billion outstanding then, $1.448 billion was secured by inventory. Wholesale and retail trade had, respectively, $573 million and $273 million.[19] The captives tabulated and presented in Table 2 held $936.7 million of wholesale notes receivable in 1955, virtually all outstanding to trade businesses. Through 1960 the captives' floor plan loans grew to $1.796 billion, but we have no compara-

---

[18] Ref. (12), Table 6, p. 405.

[19] Ref. (3), Appendix, Table 3, p. 1126.

tive figures for the banking system then. The place of inventory financing in the operations of commercial finance companies and factors is evident from the paltry $46 million of inventory financing outstanding in the aggregate $1 billion portfolio of business loans of 30 companies at the end of 1956.[20]

The 30 commercial finance companies and factors cited above had committed about 31 per cent of their assets, or $314 million, to equipment financing loans in 1956.[21] Commercial banks had equipment-secured loans of $2.2 billion outstanding in 1955,[22] but this consisted primarily of direct loans to the purchaser. Unfortunately, we cannot pinpoint the amount of equipment financing by the captives for direct comparison. The $507 million retail financing held by captives in nonelectrical machinery and $649 million in transportation equipment sans GMAC is probably almost solidly equipment sales financing of the sort under discussion. An indeterminate but substantial fraction of the more than $895 million of retail paper held by captives in electrical machinery was installment financing of equipment.

An analysis of member bank loan practices during 1955 to 1957 failed to show discrimination against small businesses either in quantitative rationing or interest rate charges.[23] Thus, rather than burgeoning captive company loans being refugee requests as a cyclical money market phenomenon, it is more plausible, but still unproved, that the inventory and equipment financing which the captives originated during this period represented a fund flow into a segment of the credit structure feebly supplied perennially.

Some incidents have come to light which are consistent with the proposition that some captives have for a time made loans not marketable elsewhere in the financial system, and that their activity led to imitative competition by other lenders which permanently (?) changed the availability of funds to these types of loans. A number of corporate financial officers have insisted to the author that they would like nothing better than to be out of the finance company business, and some have succeeded in getting out when the "need" for their captive evaporated with the adoption of more charitable (or aggressive?) attitudes by the banking community toward the type of paper generated by the parent company's operations. Several instances of liquidation are attributable to this, and, after years of expansion, at least one of the oldest and largest captives has found the growth of its assets stymied by competition of banks for eligible paper.

---

[20] Ref. (12), Table 1, p. 451. This survey covered companies with an estimated one-half of total assets of commercial finance companies and factors.

[21] Ref. (12), Table 1, p. 451.

[22] Ref. (3), Table 3, p. 1119.

[23] Ref. (1).

A particularly interesting dimension of the captives' lending is its geographical distribution. Officers of several different captives in different product lines have stated in interviews that the bulk of demands upon them for financing comes from areas where banks are thinly seeded; in states and areas where banks are plentiful and/or aggressive no paper comes to the captive. The same comment appears to apply less extensively to the independent finance companies. The captives draw funds heavily from the commercial paper and bank loan markets, usually without geographical limitation, and redistribute the funds through a national apparatus; thus, an extremely important function is performed in knitting together separate loan markets.

V. CAPTIVE FINANCE COMPANIES, ASSET AND CAPITAL STRUCTURES, AND COST OF CAPITAL

Here we consider the effects of captive finance company operation upon the asset and capital structures of parent and subsidiary, outline the moot issues of the reality or unreality of the supposed usefulness of captives as a device to lower cost of capital, and very briefly indicate the obstacles in the path of verification or denial of the captives' alleged advantages.[24]

Before proceeding it is imperative that we state precisely the financial effects of using captive finance companies. First, a liquid asset is disengaged from the asset structure of the nonfinancial parent, and a wholly owned subsidiary is formed to hold it. Second, matching the asset holdings, a capital structure is created consisting of debt in multiples of an equity base. Consequently, the proportion of debt financing in an imaginary consolidation of parent and subsidiary claims structures is raised.[25]

To pick the extreme for illustration, the ratio of all kinds of debt to equity of General Motors Corporation, not including GMAC, at December 31, 1960 was 0.347. With GMAC consolidated, the ratio is an interesting 1.15. Debt was 25.8 per cent of total assets of the deconsolidated parent and 53.5 per cent of consolidated GM and GMAC assets. All of this while the parent company has exactly one modest issue of long-term debt outstanding. In short, the creation of a legal shell for ownership of notes receivable, it would be argued by some, has enabled the parent company to parlay asset segregation into leverage ratios known only rarely in prospectively solvent nonfinancial corporations outside the utilities. The

---

[24] We disregard administrative conveniences entirely, and concentrate solely on questions of capital structure.

[25] Typical bank credit lines provide for bank borrowing to the extent of three or four times equity plus subordinated indebtedness. In practice ratios run higher or lower depending on size and age, and on factors unique to individual cases.

object in this, according to many statements received by the author in interviews, is to finance an asset stock with "cheap" debt funds.

Can the Sum of the Parts Be More Than the Whole?

Allegedly, segregation of the portfolio of receivables from other assets reduces the risk borne by creditors of the captive. A low-risk asset structure composed of liquid, determinate-life notes receivable backed by some title retention device acts as security. Thus, through creation of a prime credit shell, debt/equity ratios for the captives are achieved much closer to those typical for independent finance companies than for nonfinancial corporations.

As a matter of empirical fact there is nearly always a clean break between parent and captive with respect to liabilities. There are instances of parent guarantees, partial or full, of the subsidiary's borrowing, but they are relatively few.[26] More frequently, the quality of the subsidiary's notes is upheld through repurchase agreements maintained with the parent, though this is far from being a universal feature. Of course, the notes are - so far as I know - invariably purchased at less than full face value. Also, as a point for the reader's information, nearly all parent companies deconsolidate their active credit captives in published financial statements. The captive's statements may or may not be published.

Does the captive realize leverage advantages that are <u>really</u> unobtainable by other means? If a captive company does, in fact, enable the parent to raise its debt/equity ratio, is this supposed advantage real?

It is easy enough to cite analogous means of financing a company's receivables which provide leverage equaling that obtained through the captive device. By lending on the security of receivables to the extent, say, of 80 per cent of the face amount, a financial institution provides a debt/equity ratio for the financing of that asset of 4 to 1, fully the equal of the leverage enjoyed by most captives. Assignment of open book accounts receivable to commercial finance companies in continuing arrangements would provide leverage of the same order as a captive.

26

Most of the bank loan officers contacted by the author are of the opinion that, whether a formal parent guarantee exists or not, default by the subsidiary would not be allowed because of its deleterious effect on the parent's credit. It is uncertain whether this is true or not. Irrespective of its truth or mere wishfulness, however, bankers are discomfited by the recalcitrance of parent companies in granting guarantees. The author is at a loss to understand the fondness for parent guarantees. Barring "imprudent" management, the security of receivables is so much greater than that of the illiquid brick and mortar of the parent that its default would precede that of its subsidiary.

Moreover, it would be difficult to maintain in abstraction that the segregation of receivables in a captive offers creditors a form of security not otherwise available. The one-time pledge of receivables as collateral on, say, a 60-day bank note is closely analogous, and the assignment of open-book receivables in continuing arrangements for the advancement of revolving credit is even closer. The intrinsic security of receivables is their liquidity, and segregation for the lender's benefit by different means could perhaps differ nominally in form but not in effectiveness so long as it was legally defensible. Working capital restrictions for bondholders' protection, closely analogous in effect to receivables segregation, are an old story. In this respect the captive is not a new technique.[27]

Thus, it is not at all clear that leverage obtained in financing receivables via the captive device is significantly different and that security offered is a departure from what could be achieved by other means. When queried, corporate managements have replied, not invariably but nearly so, that although these points are true in abstraction, prejudices (unthinking?) in the lenders' world against balance sheets of nonfinancial corporations with heavy debt/equity ratios prevent attainment of the desired leverage without the separation of captive and parent company. Deconsolidation of a captive cleans up the right hand side of the parent's balance sheet.

I can only comment that deconsolidation cannot evaporate liabilities, or at least it shouldn't.

Muddy as the water is on this score, though, it seems clear in comparison with the second objection which questions the very financial goal of the captives - the use of leverage. A recent controversial theorem has it that average cost of corporate capital is a constant independent of capital structure. The premise supporting this proposition is that the characteristics of an income stream are produced by a given asset structure irrespective of the capital structure which is its financing, and that the various divisions of this income stream between classes of claimants cannot raise total market value of bonds and stock outstanding. If disparities develop between market values of levered and unlevered firms with the same basic characteristics of income, as determined by assets, a

---

[27]The validity of this point does not depend in any way upon whether or not the captives are innovative or redundant lenders. One might be tempted to, say at first, blush, for example, that banks might be willing to lend on the security of receivables held by a captive but not on the same security on their own account, and that this indicates the security form of the captive is significant. But the point is precisely that the captives differ little from conventional receivables lending in the form of security they offer. The paradox would lie really in, say, a bank making the loan in one instance and not making it in another case on essentially the same security.

profit incentive will exist for investors to divest themselves of the securities of the overvalued firm and acquire the higher rate of return securities of the undervalued one. The argument runs to the effect that the portfolio switching necessary to do this will throw overvalued securities on the market and add to the demand for undervalued ones, and equation of values will prevail in equilibrium.[28]

The view contained in this argument is very much to the point in the case of the captive companies. Since, with exceptions unimportant here, the captives are wholly owned subsidiaries, the risk shouldered by holders of the parent company's equity is precisely the same as if the receivables held by the captive had been financed and secured comparably within the financial structure of the parent.[29] Giving creditors prior claim on a prime asset will vitiate the equity holders position. The "arbitrage" process contemplated in the argument outlined above should, therefore, perform to hold the average cost of capital on a level equal to that of unlevered firms of the same risk. Incorporation of "low cost debt" in the subsidiary's claims structure, and indirectly in that of the parent, should result in a re-evaluation of the parent's total earnings stream by the equity market which will result in reduction of the market value of stock outstanding. According to the theorem, the cost of equity will rise as more debt is incorporated, precisely offsetting the effect of the cheap debt.

The importance of this argument is acute if the captives turn out to be intermediaries between risk loans and the financial markets which will not service the loans directly, or which would not have permitted the parent to debt finance the receivables through a conventional asset and capital structure. Having rejected direct association with the risks involved, the financial markets are obliged in theory to reassess their valuation of the earnings stream implicit in the earnings/price ratio of the stock of a parent, especially since the consummation of financing for the captive implies that satisfactory guarantees of fixed claims have somehow been peeled off from the parent-captive combine leaving the chaff for equity claimants.

The profession is not agreed on the point that average cost of capital is independent of the debt/equity structure. Our point here, however, is not to broach this argument but only to frame this aspect of the credit captives within recent theoretical developments and thus sharpen the issue. Whether leverage of the captives somehow betters the lot of the common stockholder is but one specie of a genus of corporate capital structure problems.

---

[28] Academic readers will recognize this summation of Modigliani and Miller in Ref. (10).

[29] Assuming the same leverage attainable with and without a subsidiary.

CONCLUSIONS

Impressive growth of inventory and installment receivables lending by captive finance companies seriously extends the role of parent corporations as financial intermediaries. Questions of both theoretical and practical significance attach to the captives' operations. Our factual knowledge of the mushrooming captives is regrettably slim, and, perforce, interpretation of their economic significance is fragile. We have, though, blocked out broad areas for inquiry.

Despite origins and marked asset growth of the captives in periods of cyclical sales expansion and monetary policy restraint, several facets of their lending imply secular improvement in the financial markets. The view from under their parents' wing enables the captives to evaluate the risk of inventory and installment receivables financing more accurately than outside lenders can. This vantage point in risk management together with opportunities to achieve economies of scale in loan administration lend credence to the hypothesis that captives improve allocational and operating efficiency of the financial markets. An adjacent question relates to the use of the leverage by the captives to finance their lending. It is perfectly imaginable that a nonfinancial parent corporation could undertake precisely the same volume of lending with or without a captive. The case for use of a captive as a fund raising device must rest on the unproved proposition that it lowers average cost of corporate capital.

To repeat an earlier point, this article is offered only as a progress report from my own research and a modest formulation of important issues surrounding the captives. If investigation is furthered, my purpose will have been served.

APPENDIX A

IDENTIFICATION OF THE SUBSIDIARIES

The identities of the captive companies dealt with in this research were obtained in a variety of ways. A small number of the subsidiaries are, of course, well known. The major effort in identifying the lesser known or unknown captives was a page-by-page search of Moody's Industrials, scanning the list of each parent company's subsidiaries for names suggestive of financial activities of the types considered here. This was the single richest source. Standard and Poor's list of subsidiaries was of some help. Financial periodicals carried news from time to time which evidenced formation of new companies. Interviews of company officers working in or in connection with the captive companies also yielded a few company names unknown to the author until the interview. For these due gratitude is expressed. It is true, however, that the information of this sort abroad in the trade is surprisingly slim.

For the largest part company administrators are aware only of captive companies operating within their respective industries. A few attempts have been made by the financial institutions to gather lists for their own purposes, but these have been for what are essentially sampling purposes and are not and were not intended to be definitive in any sense. A few years ago the Federal Reserve gathered a fairly good list of captives. A list of 66 captives was published two years ago in the Controller.[30] The list collected by the author has exceeded all those which he had discovered while working in the field.

The companies satisfy two conditions for inclusion. First, as a part of the definition laid down for the entire study,[31] the parent company market goods on more than a local level. In fact, most of them market goods nationally. Secondly, the subsidiary must have been a significant and active part of the parent company's operations. This elastic requirement implies quite accurately that subjective judgment was exercised. In most instances it is perfectly clear when a subsidiary has been active and significant, but some borderline cases exist. A certain number of subsidiaries were formally created but not put into operation, and some were active in pre-World War II days but not after. These were not included. A few subsidiaries have operated at very low levels of receivables absorption relative to the parent's sales volume, and there has been no evidence of other kinds of interest in aggrandizing the subsidiary. These were not tabulated either. On the other hand some captives operating on a very modest scale relative to the parent's volume were included because there was reason to believe that they were used aggressively, however limited their purpose. It is apparent that the number of companies used is the residual of a weeding process which started with a substantially larger figure. The reader should note that the figures of Table 1 refer to captive company births only, not to the net number of companies in operation at a given time. Some of the captives enumerated there were active for a period and then deactivated, but these were few. As a practical matter, the number of companies formed differs but little from the number currently active.

APPENDIX B

STATISTICAL SOURCES AND METHODS

Sources

The balance sheets which underlie the statistics presented in this paper were gathered from various sources. After identification of the

[30] Ref. (11).

[31] Page 27, ff. 2.

subsidiaries by the methods outlined in Appendix A a mailing of a form letter to appropriate company financial officers was made in the case of all subsidiaries not publishing balance sheets. A repeat mailing was tried in all instances of no response. Most of the parent companies of the largest captives now publish deconsolidated balance sheets in their annual reports, but in earlier years this was not the case and much of the point in mail requests related to obtaining data for these earlier years. In a number of cases, though, the parents did not and still do not publish balance sheets for their finance subsidiaries even though deconsolidated. Some of these companies readily supplied balance sheets upon request. In the instances of no responses several paths were open. Occasionally the corporate files of Dewey Library at MIT and Baker Library at Harvard yielded usable information through prospectuses or proxy statements which the companies had evidently not supplied through negligence or disregard rather than from reluctance to disclose the information. A few parent companies filed deconsolidated balance sheets for their subsidiaries as required by the laws of Massachusetts, and their balance sheets in the State House in Boston were used if otherwise unavailable. A very few balance sheets were obtained from the files of bank loan officers, who must remain anonymous but to whom the author is indebted, and in some instances fragmentary statistics derived from these sources served to cross check fragments from other sources and aided in piecing together balance sheets as described in the second paragraph below.

A final resort in the case of listed corporations lay in the requirement of the SEC that they file 10-K reports annually with the Commission. In some of these, deconsolidated balance sheets for the subsidiary are filed. A few of the balance sheets employed were garnered from this source by a visit to the public files of the SEC in Washington. In some few instances when it was impossible to obtain data from the sources named above it was feasible to construct balance sheets from significant fragments of information divulged about the subsidiary in annual reports of the parent companies, prospectuses of security issues, and proxy statements. A frequent source of this kind was the ubiquitous footnote to the parent company's balance sheet concerning the principles of consolidation. When the subsidiary's balance sheet was deconsolidated but not published, it was sometimes true that principal assets and liabilities were divulged in this footnote. With these fragments and under some assumptions it was sometimes possible to construct plausible balance sheets based upon some of the regular relationships between items which emerged from work with fully available balance sheets. The number of balance sheets pieced together in this fashion, however, is quite minor in relation to the total number employed and even more minor in the relation between the volume of assets involved and the volume aggregated.

For purposes of tabulation the captives were sorted into the Standard Industrial Classification with reference to the product criteria outlines

in the SIC manual and the major products of the parent companies. Although data are presented in this article only at the two-digit level of the SIC, they were dealt with in the author's work at the three-digit level also.

Accounting Methods and Conventions Employed

Some arbitrariness is built into the balance sheet statistics presented in Tables 2 through 4 with respect to dates involved. The Tables name only the various years as collection dates, and, with some dissimilarity in reporting practices of the subsidiaries underlying the data collection, a certain amount of cleanness is lost. Throughout the process of assembling balance sheets the author and his assistants observed a stricture, mostly a matter of convenience, of recording balance sheet figures as reported by the captives at the end of their fiscal years. Clearly, it is most convenient to do this, if it is not outright impossible to do otherwise.

In a certain number of cases balance sheets were available to the author at other dates, and conceivably more could have been obtained. However, the difficulty of doing so, the uncertainty of the outcome of such an attempt, and the dubiousness of its value led to pursuit of the practice as named. An attempt to obtain balance sheets for all subsidiaries at one date would have introduced as many ambiguities as it removed. A substantial number of subsidiaries, perhaps most, experience seasonality in the demands made upon them for loans, especially for inventory financing. The volume of retail loans outstanding probably swings in an amplitude proportionately smaller than floor planning because the duration of retail financing contracts extends over the seasonal swing and often over several. Nevertheless, interviews have confirmed the existence of substantial seasonality in loans outstanding. Balance sheet dates are chosen in many instances as they are by many nonfinancial corporations to minimize the effects of seasonal fluctuation in the amount of debt outstanding. That is, an off-season date is chosen for reporting. If the managements of the captives all chose to divulge balance sheets at a given date, asset and liability totals would be larger than those presented here by the amount of the (indeterminate) unreported seasonal demand. Since the captives do not have uniform seasonality, however, it is doubtful that the statistics would be "improved" in any sense. Instead, by leaving the balance sheets at their reported dates the Tables in this article have precisely the amount of unreported seasonality conventionally concealed by balance sheets. Most captives report as of December 31.

It is also appropriate to comment here upon some of the practices used in recording balance sheets of the captives. Notes receivable were carried gross of loss reserves, but net of unearned finance charges and deferred discounts. Almost without exception receivables were carried on the captives' balance sheet net of both loss reserves and the income

deferred and when single figures appeared for these accounts, a problem of dividing the sum between retail and wholesale paper was sometimes posed. This was resolved by dividing the figure arbitrarily in the proportion which prevails in the case of similar companies which did indicate the basis of their allocation of the loss reserves and deferred income between accounts. In a number of cases an industry pattern prevailed with surprising homogeneity. In some few instances data were not available to make it possible to carry receivables gross of either loss reserves or deferred income. In these instances receivables were carried at the net figures shown on the balance sheet available.

The catch-all account labeled Miscellaneous Assets is quite polyglot. It probably would not be worth mentioning the great miscellany contained therein except that in isolated instances of small captives most of their assets may have ended in this account. Included are term loans, mortgages, construction advances, leased assets, and receivables from the parent company.

Some lumping was required for presentation of the right hand side of the balance sheets so that it is worthwhile to recount some of the arbitrary practices adopted. The short-term borrowing accounts are straightforward, referring only to negotiated sources of funds. Accruals have the meaning attached to them in ordinary accounting practice except that accounts payable to the parent company were broken out whenever possible and carried under borrowing from parent companies. Some notes payable to the parent company specifically designated as subordinated were lumped with others not so designated and are carried together with accounts payable to the parent company under "Borrowing from Parent Cos." Dealer's reserves, whenever specified as a separate account, were carried under accruals. The major items in accruals typically were accrued interest and taxes.

Equity includes stock outstanding, with negligible amounts of preferred stock and retained earnings.

REFERENCES

1. Bach, G. L., and Huizenga, C. J. "The Differential Effects of Tight Money," American Economic Review, 52 (March, 1961).
2. Banner, Paul H. "Competition, Credit Policies, and the Captive Finance Company," Quarterly Journal of Economics, 72 (May, 1958).
3. Cagle, Caroline. "Security Pledged on Business Loans at Member Banks," Federal Reserve Bulletin, 45 (September, 1959).
4. Federal Reserve Bank of Kansas City. "Trade Credit: A Factor in the Rationing of Capital," Monthly Review, (June, 1957).
5. _____. "Forces Behind the Growth in Trade Credit," Monthly Review, (October, 1959).

6. Goldsmith, Raymond W. Financial Intermediaries in the American Economy Since 1900. Princeton: Princeton University Press, 1958.

7. Jacoby, Neil and Saulnier, Raymond. Business Financing and Banking. New York: National Bureau of Economic Research, 1947.

8. Lavington, F. L. The English Capital Market, 2nd edition. London: Methuen and Co., Ltd., 1929.

9. Meltzer, Allen H. "Mercantile Credit, Monetary Policy, and Size of Firms," Review of Economics and Statistics, 42 (November, 1960).

10. Modigliani and Miller. "The Cost of Capital, Corporation Finance and the Theory of Investment," American Economic Review, 48 (June, 1958).

11. Seed, Allen A. "Should You Set Up a Finance Subsidiary?" The Controller, 27 (November, 1959).

12. U.S. Congress. Financing Small Business, Report to the Committees on Banking and Currency and the Select Committees on Small Business, by the Federal Reserve System, Parts 1 and 2, 85th Congress, 2nd Session.

Edward H. Bowman
School of Industrial
Management,
Massachusetts Institute
of Technology

Management Decision Making:

Some Research[1]

This paper sets forth some research results concerning management decision making and a "New Theory"[2] which was derived from them. The research evolving over a period of several years has dealt essentially with the management problem of production and employment scheduling in a number of manufacturing firms. That is, how does (should) a management adjust its manufacturing organization to some facets of its changing environment? The research appears to have developed notions of a more general nature than their production problem setting, and these more general notions of management decision making are described here. The new theory and its components now lead to more questions - questions of validity, generality and operationality.

In order to introduce the theory very briefly first, for the purpose of raising the points which this paper attempts to explain in succeeding pages, the summary paraphrased from a previous paper is given:[3]
(a) In their decision making behavior, managers and/or their organizations can be conceived of as decision rule coefficient estimators (not that they explicitly are coefficient estimators).
(b) It is the variance in their behavior rather than the bias that hurts, due to dish-shaped criteria surfaces.
(c) A decision rule with mean coefficients estimated from management's behavior should be better than actual performance.
(d) It may be better than a rule with coefficients supplied by analysis.
(e) Systematic and comparative studies using this idea may lead to further ideas.

---

[1]This research was in part sponsored by the M.I.T. Operations Research Center contract with the Army Office of Ordnance Research (DA-19-020-ORD-2684) and the M.I.T. School of Industrial Management Sloan Research Fund.

[2]The research itself is described in Ref. (1).

[3]Ref. (1).

Much of the present focus of today's interest in management is on decision making. Broadly defined, this includes the activities of scanning the environment, internal as well as external to the organization, in order to determine when and where decisions are required, as well as laying out the alternative courses of action, and choosing from among them.

A distinction can be made between programmed decisions and non-programmed decisions.[4] Programmed decisions are those which are repeatedly made and for which a systematic and somewhat standardized procedure has been worked out. Determination of the weekly production schedule would be an example of such a decision. It is with this type of decision that it is most natural to conceive of a decision rule, where the results are derived in a given way from a fixed set of variables. It is in this type of decision problem that operations research and the mathematical techniques related to it have had their more common application. The procedure employed is briefly as follows:[5]

(1) Pick the measure of effectiveness.
(2) Determine the important variables.
(3) Determine the relationship of these variables to the measure of effectiveness, i.e., build the model. Probability may or may not be a major factor.
(4) Determine the values of the parameters. This determination may be a statistical or a cost accounting and/or engineering approach.
(5) Test the model. Steps (2) through (5) are repeated until a satisfactory model is constructed.
(6) Manipulate the model for purposes of determining a useful decision rule.

The non-programmed decisions are the decisions made infrequently or only once. The decision process is ad hoc and not well structured. Programmed and non-programmed are of course only convenient ends of the spectrum. The original research leading to the main ideas presented in this paper dealt with production scheduling in the sense of a programmed decision where it was natural to conceive of a decision rule with variables and coefficients. This research is described in the earlier "New Theory" paper. The example to be used here to carry the burden of explanation and illustration of the new theory is much closer to the non-programmed end of the spectrum; it deals with the problem of the proper size of branch plant operations. The objective is to illustrate the theory, rather than fully develop the particular problem.

---

[4] Ref. (6).

[5] Ref. (3), p. 294.

PLANT SIZE

Several years ago a study was made of the economics of scale of operations of a system of ice cream branch plants.[6] The company in question was operating ten plants over a seven-state area. The question asked was: what is the optimum size company plant to manufacture and distribute ice cream, given their technology? That is, based on the total sales of the company, about how many plants should be operating over the seven-state area? Though this decision was basically an unstructured or non-programmed one as far as the company management was concerned, our analysis was done in the spirit of operations research. That is, a measure of effectiveness was chosen (cost per gallon), the variables in the situation which seemed to influence this measure were identified, and a mathematical model which related these variables to the measure of effectiveness was constructed. Using the actual costs incurred by the company in operating those ten different plants under their varying conditions, it was possible to supply the particular parameters (dimensions) which related the variables to the effectiveness measure. The mathematical model was then manipulated (a derivative was taken and equated to zero) to provide the decision rule,

$$(1) \qquad V_{opt} = (K)^{1/3}\left(\frac{2b}{c}\right)^{2/3}$$

where V_{opt} is the optimum value, K is the sales density of the particular area expressed in gallons sold per square mile of territory per year, and b and c were parameters of the original model derived by analysis of the cost experience of the firm.

When the actual volumes of each of the ten plants of the firm were compared to their optimum volumes derived from this decision rule, one plant was about 50 per cent too large, and the others were too small. Most of them were about half their optimum size and several were substantially smaller. It appeared that the company was paying a premium of approximately $207,000 each year for operating these uneconomic size plants.

Now using the results of the research on managerial decision making and the new theory briefly described at the beginning of this paper, it was possible to use the management's own decision making behavior in order to estimate the coefficient in the decision rule (1). In the rule previously set down, the expression $(2b/c)^{2/3}$ is really a number. It is possible to substitute a symbol (d_1) for this coefficient which gives the decision rule

$$(2) \qquad V = (K)^{1/3} d_1 .$$

---

[6] Ref. (2).

Notice V rather than V_{opt} is used as the more general decision rule no longer assuring an optimum volume. It is now possible, by least squares statistical regression analysis, to estimate the value of the coefficient d_1 from the organization's own arrangement (i.e., decision results). In more simple terms, it is possible to plot the actual V versus $(K)^{1/3}$ as in Figure 1 (the figure here is illustrative and does not present the real data).

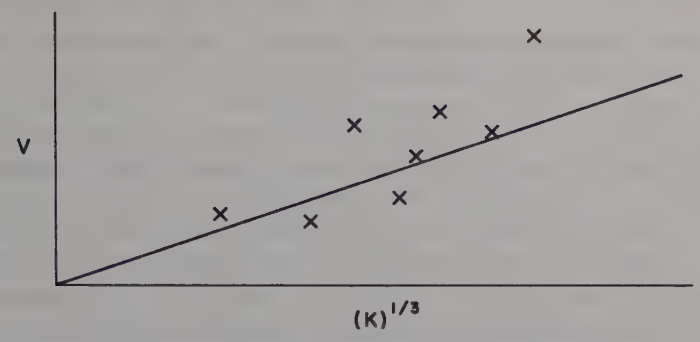


Figure 1

Because there is some variance around the line representing d_1, the variable $(K)^{1/3}$ does not fully explain the variations in V. However, if equation (2) is used as a decision rule (with the variance around the mean estimate of d_1 eliminated), the restructured system of plants would have incurred a cost premium of only $72,000 per year compared to the "optimum," which is a $133,000 saving from the $207,000 cost premium incurred by the actual arrangement. That is, using the same form of the decision rule as that provided by the analysis, but using management's own behavior to estimate the decision rule coefficient rather than that from the cost analysis, much of the cost savings ($133,000 versus $207,000) might be made available. A sketch of this idea, Figure 2, might make it clearer (again using illustrative data).

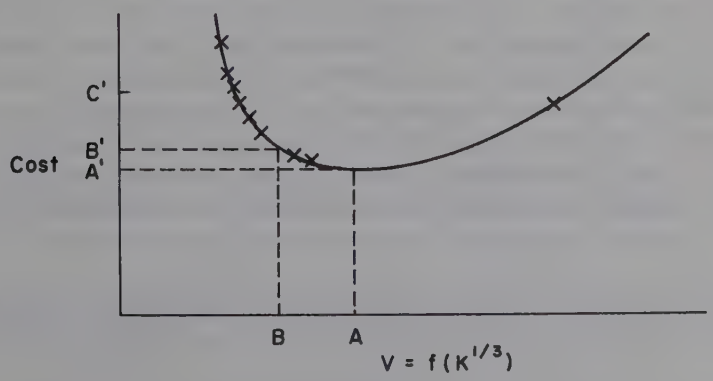


Figure 2

The volume labelled A in the sketch is the optimum (least cost) volume arrangement that would save $207,000 compared to all of the individual x's which represent the actual individual plants. The volume labelled B is the average (mean) of the present volume arrangements (not merely $\overline{V}$, as the function of $K^{1/3}$ makes this a whole family of curves) which would save $133,000. The difference along the volume axis between A and B might be labelled the bias in management's behavior, which generates a cost premium of $72,000 (B' – A'). The sum of the differences between the individual x's and B which yields the $133,000 (C' – B'), might be labelled the variance in management's behavior.

Departures of the decision making behavior of management from the preferred results then for our purposes can be divided or factored into two components, one of which, in the manner of a grand average departing from some preferred figure, we call bias (which causes a relatively small criteria loss due to the dish-shaped bottom of the criteria surface), and one of which, representing individual occurrences of experience departing from the grand average, we call variance (which causes larger criteria losses due to the individual occurrences up the sides of the criteria dish-shaped surface). It is the latter, and more important, component which seems to offer the tempting possibility of elimination through the use of decision rules incorporating coefficients derived from management's own recurrent behavior.

In the research mentioned on production and employment scheduling in four different firms, seven decision rule coefficients were estimated from much behavior for each firm rather than the one (d_1) as described here. Then the company scheduling was simulated using the decision rules containing management's own average coefficients (with the variance eliminated). The results in three of the four cases would have been less costly than actual behavior, while in two of the cases the results would have been even less costly than a rather involved operations analysis would have allowed.[7] In other words, there seem to be cases where decision rules with coefficients estimated from management's own behavior not only would allow better results than those actually incurred but also better than those provided by operations analysis. This possibility can be explained in part by the limitations of an analytic procedure. In the production and employment scheduling problem, explicit quadratic forms were used in a number of situations to represent empirical cost curves. Precision was knowingly sacrificed by this representation so that the mathematical procedure would be feasible. Then, because it is the model which is optimized, the decision rule recommended by the analysis contains this imprecision.

There is a second source of possible advantage for the decision rule (and its coefficients) derived from the study and statistical regression

---

[7] For a detailed description of this type of operations analysis, see Ref. (4).

of management behavior, compared to a rule and its coefficients both derived by operations analysis. Management may respond (sensibly) to criteria which are particularly difficult for the analyst to understand and measure. The behavior regression approach may permit the fusion of specialized competences - those of the analyst with those of the manager.

FURTHER RESEARCH

Many questions still remain unanswered by the work done so far, and much research remains to be accomplished. Questions of truth, scope, and usefulness must be answered if these ideas are to benefit anyone. It seems to fall within the province of the social sciences to consider statements such as the following:[8]

(a) Experienced managers are quite aware of and sensitive to the criteria of a system.

(b) Experienced managers are aware of the system variables which influence the criteria.

(c) Managers, in their present position through a process of natural screening, make decisions, i.e., implicitly operate decision rules, with a sense and intuition which relates the variables to the criteria imperfectly - but more erratic than biased.

(d) Most cost or criteria surfaces are shallow, dish-shaped at the bottom (top) and, even with bias in the manager's "estimation," it is the far out (variance) examples of behavior which are really expensive or damaging.

(e) If managers' behavior had paralleled the decision rules with their average or mean coefficients, their experience would have been better according to the (their) criteria.

(f) This idea also pertains to the future with the system maintaining its same basic structure and parameters.

It seems useful to attempt an explanation of why decision rules derived from management's own average behavior might yield better results than the aggregate behavior itself. Man seems to respond to selective cues in his environment - particular things seem to catch his attention at times (the last telephone call), while at other times it is a different set of stimuli. Not only is this selective cueing the case, but a threshold concept seems to apply. He may respond not at all up to some point and then overrespond beyond that. It is this type of behavior which helps explain the variance in the organization's (or its management's) behavior. A thorough review of the findings of psychology in this area may help support these ideas. Laboratory experiments where the environment is especially structured in a simplified setting should help explore these characteristics.

The idea of dish-shaped criteria surfaces is a surprisingly general one. Many different operations research type studies as well as economic

---

[8] Ref. (1).

analysis seem to show this. For instance, from the literature of econometrics reporting a substantial amount of empirical work comes the following statements:[9]

> In most cases, the penalty associated with moderate deviations of the actual instrument values (the decision coefficients) from the maximizing ones is small . . . However, if it is true that moderate deviations of instrument values from the maximizing ones are not very serious, then a more fortunate conclusion is possible in the normative sphere; viz. that approximations are much more acceptable here than in the analytical sphere. It is conceivable that in many cases, this fact can be fruitfully used in order to arrive at a certain degree of 'automation' . . .

Not only must more work be done to investigate these criteria surfaces in an empirical sense, but more investigation of a mathematical nature can be done to determine under what patterns of behavior variance the theory's results hold true.

If these ideas are true, and the "New Theory" paper presented six or seven empirical test cases, then the practical use of the ideas gets into the general areas of heuristic problem solving and artificial intelligence.[10] In a sense these are man-made substitutes and short cuts for man's own intelligence, where statistical methodology, programs, and computers (in large problems like the production and employment scheduling problem with seven unknown coefficients to be estimated simultaneously) are used to represent the relevant characteristics of the decision process.

One of the major problems in explicitly using our heuristic approach is to provide for adjusting or learning mechanisms as the environment changes. As problems unfold over time, in order to use past experience (regression of behavior) to solve current problems, a time lag is inevitable. So it seems that we must be concerned with the general ideas of feedback. This presents the possibility of studying and using not only man's reaction to his environment, but also the changes in the structure of his reactions as the structure of the environment changes.

If a manager is to use these ideas, it presents the advantage of turning his focus of attention from the decision to the decision process. This seems to offer the possibility, in part, of moving away from putting out fires to the more professional task of designing intellectual processes.

Probably the real test of the "New Theory" will come as real organizations attempt to implement the ideas. A large "technologically ad-

---

[9] Ref. (7), underlining supplied here.

[10] See Refs. (6) and (5).

vanced" manufacturer is now developing decision rules with coefficients derived from management's own behavior for use in several problem areas. The areas so far fall into the programmed decision category. As experience is gained from research into the heuristics of non-programmed decision making which Professor Simon is reporting, it may be possible to abstract some of these decision processes into a common program permitting further generalization from experience. That is, these more abstract decision models may require parameters or coefficients which can be supplied from experienced management's behavior.

REFERENCES

(1) Bowman, E.H. "A New Theory About Managerial Decision Making," paper given at the Conference of Factory Scheduling at Carnegie Institute of Technology's Graduate School of Industrial Administration, May 12, 1961 (to be published).

(2) Bowman, E. H. "Scale of Operations - An Empirical Study," Journal of the Operations Research Society of America (May - June 1958).

(3) Bowman, E. H. and R. B. Fetter. Analysis for Production Management. Revised Edition. Homewood, Illinois: R. D. Irwin, 1961.

(4) Holt, Modigliani, Muth and Simon. Planning Production, Inventories and Work Force. Englewood Cliffs, New Jersey: Prentice-Hall, 1960.

(5) Minsky, M. "Steps Toward Artificial Intelligence," Proceedings of the I.R.E., January 1961.

(6) Simon, Herbert. The New Science of Management Decision. New York: Harper & Brothers, 1961.

(7) Theil, H. Economic Forecasts and Policy. Amsterdam: North-Holland Publishing Co., 1958.

Chadwick J. Haberstroh
School of Industrial
Management,
Massachusetts Institute
of Technology

Organization Structure:

Social and Technical Elements

In discussing the art of organization, as with most other topics dealt with by students of business and practicing managers, the emphasis is on finding ways to help do a better job, whatever the job may be. At times discussion of organization gets around to the question of how the job to be done is chosen. This question usually concerns philosophers and statesmen more than it does students and practitioners of management. This paper will concern a narrower question that lies somewhere between these two, namely, what aspects of organization structure are essential to the technical achievement of the job to be done and what aspects are given by social considerations extraneous to the task. By organization structure I mean the institutional arrangements that are consciously prescribed or at least accepted by the formal authority of the organization.

It is only reasonable to ask this question in the light of a great deal that has gone before. Students of organization and management are now able to present convincing evidence that certain ways of organizing give greater efficiency to the performance of certain kinds of tasks. We do not yet have the whole truth, but we do have some of it. A conservative application of this kind of knowledge requires not only that we be able to affirm the benefits of the new technique, but also that we are able to establish the dispensability of the institutional arrangements that are going to be replaced. To establish this requires study of the origin and functions of existing structures.

The same history that produced the possibility of phrasing the question and the need for answering it has also provided a considerable amount of research that bears upon the answer. This research builds up our knowledge of society fully as much as it does upon our knowledge of business or public administration. Furthermore, the interaction of social and technical effects on organization structure can best be studied only if the social component shows some variation; and this places especially great emphasis on cross-cultural studies. It is therefore no coincidence that most of the work to be cited below, as well as my own interest, stems from consideration of similarities and differences between organizations in different socio-cultural environments.

THE ORGANIZATION IN A CHANGING SOCIETY

The foundations of this line of research were laid down a generation or two ago by the sociologists Weber, Tönnies, and Maine. The major themes of their work are (1) the contrast and conflict between a traditional, person-oriented type of social organization and a rationalized, impersonal, task-oriented type and (2) the transition in the course of the social and economic development of Western society from the former to the latter. Before dealing with the differences, it is well to analyze certain aspects of structure which seem to persist from relatively primitive organizations in traditional societies to the most complex of our modern bureaucracies.

The first of these is the relatively great importance of the organization head, whether he is king, chief, president, or chairman. He seems invariably to perform a major role in the definition of the organization's character. The second aspect is the principle of hierarchy, the maintenance of definite status levels among the participants in the organization implying degrees of worth. The third aspect is the attention to processes of making the lower echelons responsive and tractable in the service of the interests represented at the top. I do not wish to suggest that these are immutable principles of administration, but only that they are empirical regularities in the mainstream of organizational evolution as we have observed it to date. An important qualification, however, is the substantial weakening in degree of each of these features where the influence of authoritarian values has declined.

Now let us turn to the dimensions along which major changes have occurred and explore the implications of these changes for the arts of organization and management. First, the man at the top of the organization changes during this process from the master whose status is guaranteed by inheritance of title, land, or wealth, to the functionary whose qualification for office is his ability to achieve. His position is not given by custom or tradition, but rather by some form of examination. In the same way, the relationship between hierarchical levels changes from that of master-servant to that of supervision of one office by another. In modern organizations, authority rests in the office held and has a defined and limited jurisdiction over subordinate offices only, not the persons in them.

The second dimension is the change in emphasis from the personal rights and welfare of the individuals concerned to some impersonal task to which all are equally committed. In traditional society, the chief aim of an organization is the satisfaction of the personal desires of the master. In return, the servants are guaranteed certain rights and privileges. The tasks they work on, however, are given almost entirely by custom and tradition. In the newer organization, the personal welfare of the participants is secondary. The emphasis in the actual day-to-day functioning of the organization is on an impersonally determined task to which personal

interests are either artificially related, as in the wage bargain, or only remotely connected, as in the interest we all have in a stable and productive economy. Thus, when authority is exercised, the decision takes into consideration primarily the implementation of task goals, rather than any personal preferences. Under the traditional system, the goodness or worth of an organization could be judged directly in terms of the welfare or preferences of the participants. In modern organization theory, we need to add the dimension of effectiveness in achieving the impersonal purpose to which the organization is committed.

There is another dimension that perhaps underlies both of those discussed above and extends more broadly to every aspect of organizational functioning. This third dimension for change is change itself: the degree and rapidity of innovation with which the organization must cope. In traditional society it was a rare generation that could not look forward to living essentially the same life as its immediate ancestors lived. In modern society, few people can slip quietly into grooves clearly defined and made smooth by their forebears. These patterns are swept away by outside forces, among which is an unceasing quest for increased technical efficiency. The quest for technical efficiency, in turn, is enforced by the existence of organized commitments to an impersonal task.

Even the most traditionally oriented societies of today, however, are not immune to change. Developments in North America and Europe spread to create crises in the most remote parts of South America, Africa, and Asia. What is lacking in the more traditional societies is the capacity to cope with change, and that deficiency is mainly in the organizational forms they have developed and in the cultural values those forms reflect. One simple example of this which has developed in New Guinea is the "cargo cult." Natives have reacted to European encroachment by forming cults whose mission is to bring them the material advantages of the whites by supernatural means. They thus try to adapt to change within their standard technology and organizational form, but with no chance of success. However, even in the most advanced societies, the magnitude of present-day change is difficult to cope with. Our own managers may be making the same type of mistake less obviously.

As mentioned previously, the most important elements carried over from the traditional position of the master to that of the modern administrator are the central importance of his own role, his superordinate position in hierarchical relationships, and his need to control the actions of subordinates. More briefly stated, he inherits the power position of the former privileged classes. But whereas the power of the master was used for his own personal ends, the power of the administrator is used for the accomplishment of the impersonal task. One does not normally find a complete transition even in so-called "developed" societies. Incumbency in a high-level administrative position confers high social status and in-

creased personal welfare, and when a high-level position is to be filled, an effort is made to find a person whose present social status is nearly equivalent. But the difference in emphasis is clear enough, and the trend may be expected to continue even in "developed" nations as ideological forces work themselves out.

The most important feature of administrators' power is their control over the affairs of the organization as a whole(10). The primary focus is in the choice of the organization's purposes, those actions that give the organization a unified, coherent character. This is mainly an aspect of the organization's relationship with the rest of society. It is in this phase of its existence that the organization has an impact on the affairs of its time.

The policy choices involved may have consequences for the internal equilibrium of the organization. In particular, certain of them may be crucial for its continued survival. The need to assure the survival and continued good health of the organization will be an important influence in administrative policy. The need is hard to separate from the personal desires of the officials for status, power, and remuneration. In a person-oriented cultural matrix, one would expect to find relatively more expression of the latter need and less concern for the organization as such or as a means for achievement of the impersonal purpose. The fulfillment of these internal needs may require adaptation, the act of seeking for a new equilibrium with a changing environment, even including major realignment of the impersonal purposes in the interest of preserving the organization. In this way, the social instrument may itself create forces that influence the ends for which it is used. The aims of organized power have thus become depersonalized, subject to contention and decision and accordingly greatly increasing the range of alternatives open to a society.

Among the impersonal ends of modern organizations we can find many that correspond to social ideals, as contrasted to material interests. A corporation may be dedicated to improving the welfare of its work force or to promoting a "free-enterprise" ideology as well as to profit-making. It could work toward all these ends and at the same time maintain a policy of nondiscrimination in employment between races, a valuable service to the ideal of racial equality. These possibilities arise because of the use of formal rules as a technique of administration. Once the practice of administration by rule is established, organization power can be mobilized around whatever ends are chosen by the administrators.

The techniques of holding power (or of administration, if one prefers) have changed as much as the objective of its use. The functionary does not rule by virtue of personal position or property, nor by force of personality. His central instrument of power is the creation and maintenance of the fictions of organizational authority and jurisdiction. In mod-

ern organization these are made the subject of conscious and rational planning, not inherited as in traditional social structures. There is a limitation, however, to the power position of the administrator. If he is not aware of the power position of others as well as his own, his organization will founder in opposition and conflict. The organization's purpose and its means of attainment must be so chosen that the different elements of the organization can unite in their realization. In addition, they must be so chosen that they will not run afoul of external and superior political or economic power.

The nature of organizational authority has been discussed extensively in the literature. It is generally agreed that authority owes its force not to the command of him who holds the authority but rather to the obedience of those who accept it. Thus we must seek the origins of authority in the feelings of legitimacy held by the subordinates.

This was a major starting point of Max Weber's analysis of social organization[3]. Weber, however, devoted his efforts mainly to exploring the social consequences of variation in the culturally given source of authority, i.e. whether the people perceive legitimate authority to lie with the emergent leader, with the family, with clerical or secular organization, with traditional or legal government. His closest approach to discussing the origins of these perceptions of legitimacy is the concept of "charisma." This is the relation of dependence and subservience that is observed between the emergent leader and his followers. Weber hypothesized that this could be routinized by transference to the leader's family or to an organization founded by him. This process could begin or reinvigorate a traditional system of government such as patriarchialism or feudalism. The development of a legal, bureaucratic system of governance was more a new and flexible method of employing existing authority (that of religion, sovereignty or property) than of any new concept of its origin.

Although modern society has long been accustomed to the legal processes of delegation of authority and definition of jurisdiction, it is only recently that the question of the origins of the sense of legitimacy has come under scientific scrutiny. Although charismatic leaders seem always to have used doctrine to win and inspire their followers, a psychological theory of how the process of influence takes place is a recent development[9].

The result of this for organizations is that we now have not only the legal technique of manipulating the perceptions of legitimacy of those who are already convinced, but also a scientifically-based technology for doing the convincing, for creating the sense of legitimacy in the first place.

With this new technology available and commonly understood, the traditional sources of authority (sovereignty, religion, property) can no

longer be taken for granted. The words "lobbying" and "subversion" are two of the most common words in our vocabulary for discussing government, and they do not connote obedience to or respect for the legitimacy of sovereign authority. The legitimacy bestowed by property ownership cannot realistically be taken as the basis of authority in the modern corporate business community, although it remains so legally(5).

As the tempo of economic, social, and political change increases, organizations will find it increasingly necessary to adapt not only their policy but also very fundamental aspects of their institutional structure in order to continue operating successfully. It remains to identify a few common aspects of organization structure and explore their social and technical functions before discussing some possible changes.

SOME STRUCTURAL ELEMENTS AND THEIR FUNCTIONS

The preceding section has, by exploring the natural origins of some basic aspects of organization, laid the foundation for the assumption that organization structure can be engineered to realize specific purposes. The question of whether a given element of structure serves social or technical purposes resolves into the question of the function of that element in the total picture. The technical elements are those that are functional to the execution of the work ostensibly being performed. Elements are social to the extent that their role is determined by the social milieu. A social element in a given organization may be tailored to the self interest of an individual person, family, or elite; or it may be an institutionalized element of the culture as a whole.

There is probably no better or more generally appreciated example of the difference between social and technical elements than the sociological distinction between ascription and achievement. In the context of organizations, these processes appear most clearly in the distribution of rewards and in the recruitment of members. Rewards and positions are given by ascription when they depend on the class or family background of the participant, and by achievement if they depend on performance. Ascription is a broad enough term to cover such extremes as nepotism in public service and the inheritance of a family business, as well as many in-between areas where the question of public approbation is less determinate. In our stereotype of traditional society, virtually all positions and rewards are determined by ascription. In modern society achievement is the dominant ethic for both recruitment and reward in organizations, although Mills(6) advances the thesis that the really large payoffs are made and all organizational positions above a certain level are filled by ascription from a "power elite."

Presthus' recent research on work organizations in the Near East illustrates through institutional analysis the social basis of ascriptive re-

cruitment and payments as well as other organizational practices to be discussed below[8]. Recent studies by Marsh[4] and Trow[11] are typical in showing the intermixture of ascriptive and achievement-oriented recruiting processes in eighteenth and nineteenth century Chinese government and in twentieth century American small business respectively. Their method was to study the circumstances of a sample of appointments and promotions. Marsh related the highest rank gained to seniority and entrance examination as achievement factors and to the purchase of office and family background as ascription factors. His conclusion was that the latter factors were slightly more important. In a sample of small U. S. companies, Trow found that a family heir was the first choice for succession to the top post. Promotion or outside recruitment were last resorts. The heir's objective qualification, however, was an important determinant of the manner and timing of his succession and of the later profitability of the company.

Almost anyone will accord the task-oriented organization, that recruits and rewards according to achievement, technical superiority over the person-centered organization that uses ascription. The evidence for this proposition, however, although sound enough, is global and impressionistic in nature. Although some environmentalists would blame it on the contiguity of coal, iron ore, and limestone deposits, the best evidence is the wide disparity in national income between the underdeveloped, traditionally organized societies and the modern, rationally organized ones - and not the least important datum is the rapidity with which Russia was able to change from being one of the former to one of the latter through a purely organizational revision.

The evidence for the technical utility of more microscopic variations in organizational structure is also mainly impressionistic because of the lack of a generally applicable standard of technical effectiveness. This does not mean that the evidence we have is useless, however, and I will attempt to employ it to evaluate certain important aspects of organizational structure. These are the use of hierarchy, the basis for specialization, the use of auxiliary labor as opposed to an autonomous workforce, and the use of authority over property.

The use of a hierarchy of authority is almost universal in complex organizations, and it plays a part in most of the management topics that are current in the literature. These topics include decentralization, participative management, and executive development. Not the least of them is the whole modern emphasis on management as a function, a kind of work[2]. Much of the discussion of these topics has in one way or another focused on the dysfunctions of the hierarchical principle and ways of overcoming them in order to realize maximum organizational effectiveness. For example, ideas of decentralization and delegation involve an expansion of executive resources by locating ultimate decisions at lower and more

numerous points in the hierarchy. Participative management also attempts to enlist creative talent from lower ranks, but without entirely giving up jurisdiction by the upper layers. All of these techniques also lighten the psychological burden of authority by reducing the degree of domination of the upper levels over the lower. Executive development and other topics in the area of managerial functions refer directly to rationalizing the use of hierarchy in the service of the organization goals by exploring the contribution of managerial personnel to task achievement. Executive development concerns rational planning for the continuing availability of competent managers at each hierarchical level. That this may be functional for the task is shown, for example, by Trow's work(11) which demonstrates a relationship between conscious planning for executive succession (whether of a family member or not) and later profitability of the organization. All of the developments mentioned weaken as well as modify the traditional principle of hierarchical authority, either by reducing hierarchic domination or by proposing objective tests of performance for candidates for these positions.

Sociological studies place this trend in a broader perspective. Recent work by Udy explored both social and technical aspects of hierarchy, as well as other organizational characteristics. His method was to analyze a sample of specific productive organizations from different non-industrial societies. The data concerning each organization were coded from anthropological sources; the technique for analysis of the data was to apply simple statistical tests of relationship between social and technical factors on the one hand and organizational factors on the other. One such investigation(12) shows an inverse relationship between characteristics that correspond to the usual conception of hierarchy and characteristics of rational, task-oriented organization. The first group of characteristics includes three or more levels of authority, non-working members, and rewards differentiated by office. The second group includes specific limited objectives, rewards dependent on achievement, a contractual type of recruitment, and compensation paid by the higher echelons to the lower. This result implies that in a diversity of cultural settings people faced with the same elementary productive problems have found hierarchy generally incompatible with rational task-oriented organization.

Udy's latest work has also given some evidence for a social origin of hierarchy in the form of an association between "custodial" organization (recruited by political authority) and the number of levels of hierarchy[1]. But Udy's work also shows another side to hierarchy, namely a relationship between a technical factor, complexity of process, and the presence of hierarchy(14).

---

[1] Personal communication.

A reconciliation of these findings is that a certain amount of hierarchy is functional to the task in facilitating coordination, but that hierarchy tends to become infused with social values that may exaggerate its importance or modify its operation. Thus, for example, in Presthus' study of a Turkish coal administration[8], the division between upper level posts and middle and lower posts tended to correspond to strata of social class. Similarly, Udy has found some evidence of custodial chiefs generating organization rules that serve to defend their organizational position. He also found examples of overextended hierarchies generating new work that would justify their hierarchical position. Both of these tendencies are consistent with a great deal of previous thought and experience. The reader might consider, for example, Mosca on juridical defense, or Parkinson's Law.

The idea of specialization has been touted at least since Adam Smith as enabling the kind of technical progress that leads to social wealth. This concept refers primarily to horizontal specialization as between different kinds of craftsmanship. A large part of organizational thought has been devoted to the invention of effective bases of specialization, for example place, time, persons, things, and processes[2]. The notion of managerial functions opens up the question of vertical specialization, and the manager has been presented in turn as a planner, a coordinator, a specialist in interpersonal relations, a personnel recruiter, a communication link, and a blesser of proposals. The traditional specialties of the upper social classes, however, have been less varied and can be grouped rather easily under the categories of dominance, ownership, and leisure. These social factors are certainly not independent of organizational effectiveness, however. Dominance and control over resources are common and obvious ways of maintaining coordination, although they just as obviously are not always so used. Time to think has always been mistaken for leisure, although leisure is not always accompanied by thought.

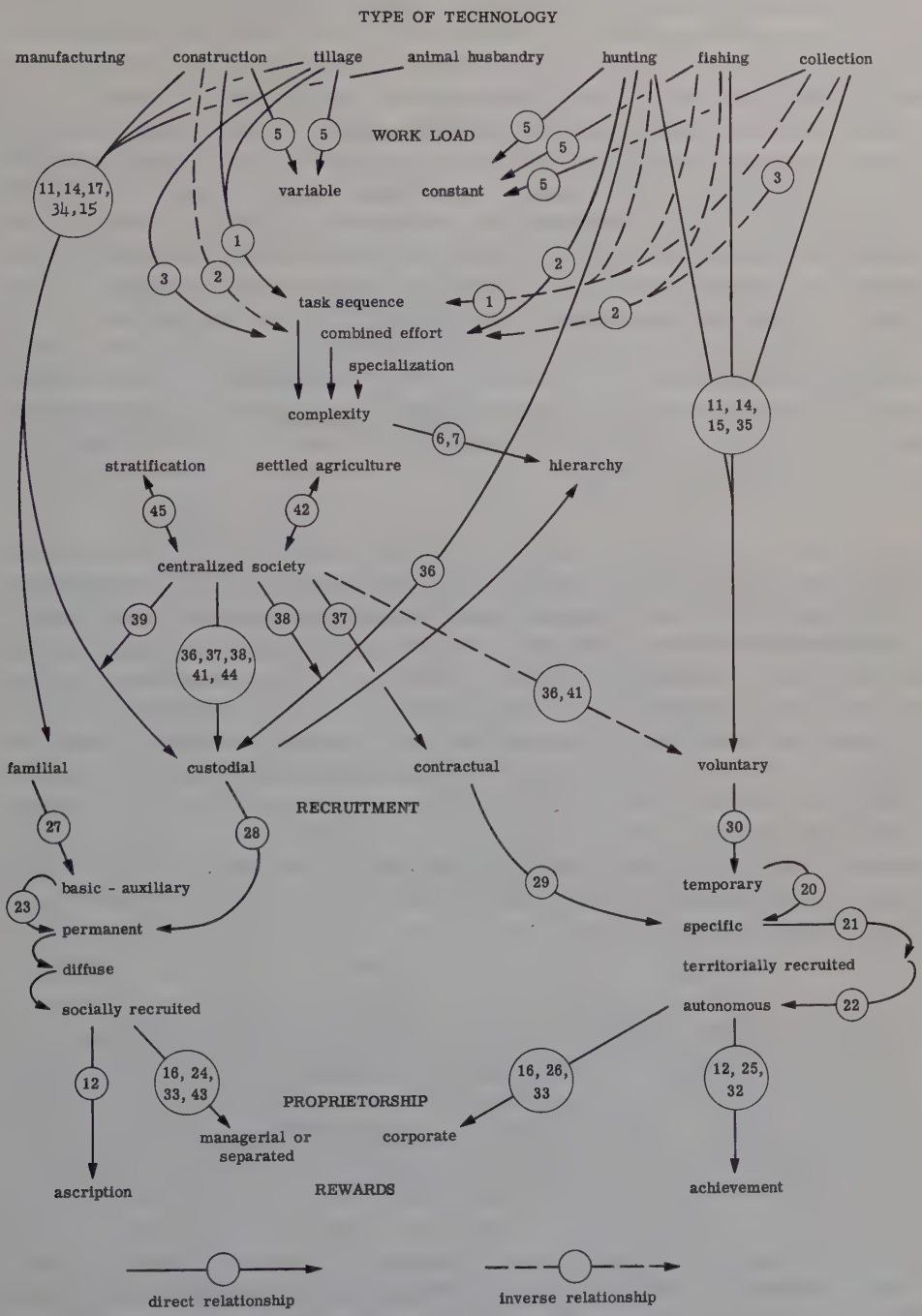
Modern trends tend to make many of the managerial functions horizontal rather than vertical specialties. It is probably also true that this can be carried much further than it has been, and many writers have emphasized the use of hierarchy to conceal dysfunctional dominance, ownership, and leisure. The result of this line of thought is a presumption in favor of the flat organization, although it cannot be said to justify the complete abolition of hierarchical relationships.

Another interesting sidelight on specialization comes from one of Udy's results[13]. In the portion of his sample where data were adequate, there was a definite scale relative to specialization. Virtually all organizations that had developed specialization of labor did so in the presence of

---

[2] Ref. (2), p. 128.

DIAGRAM OF IMMEDIATE EFFECTS*



* Adapted from S. H. Udy, Jr., Organization of Work. Numbers refer to hypotheses listed therein.

a system of compensation of workers by managers. Beyond this, the presence of specialization in work was a uniform precondition for the payment of rewards according to achievement criteria. In turn, compensation, specialization, and reward by achievement all seemed to be necessary for the development of a basically contractual system of recruitment. This finding may be relevant to problems of economic development in societies where compensation and specialization are well developed, but where modern organizational techniques are compromised by absence of the institution of reward according to achievement(7). This tendency was evident in Presthus' study of the Turkish coal administration and in my unpublished studies of various Latin American work organizations. One might compare the Russian case, in that whereas the Czarists had developed systems of compensation and specialization, the Communists added material achievement as a dominant ethical consideration. Communism does not noticeably emphasize participation on a limited contractual basis, but some recent events (i.e. increased permissiveness toward mobility of labor) tend in this direction.

As can be seen from the diagram of immediate effects that I have constructed out of Udy's findings, his pre-industrial work organizations tend to polarize into two classes: the traditional bureaucracy, permanently organized, with diffuse ends, recruited on the basis of prior membership in other social organizations, and with ascriptive rewards; and the rational, task-oriented organization, having specific limited material objectives, recruited on the basis of an equal chance of participation for all available persons, having an autonomous work force, and emphasizing achievement in the distribution of rewards. One of Udy's main conclusions is the paradox that economic and social development from primitive society tends to lead into traditional bureaucracies rather than into the rational, task-oriented organizations that can contribute to further substantial progress.

It is also interesting to note that the contractually recruited organizations in Udy's sample tend to be of the rational, task-oriented type, whereas an organization that differentiates its members into a permanent cadre plus an auxiliary, temporary work force tends to be a traditional bureaucracy. It is easy to spot the exploitative element in the use of an auxiliary work force, as for example in the distinction in many of our own industrial organizations between the permanent management cadre and the expendable hourly-paid work force. This idea is nicely illustrated in Presthus' study where there was a large auxiliary labor element employed at subsistence wages and under conditions only slightly removed from serfdom. This, of course, results in low productivity, and this is one ground for recommending employee security as a factor contributing to organizational effectiveness. More relevant in modern circumstances where security has been realized are the great gains in effectiveness ex-

perienced by some companies that have clasped their labor force more closely to them by such devices as profit-sharing and participative management[3].

One might similarly speculate about Udy's finding that traditional bureaucracies tend to rely on managerial or absentee proprietorship, whereas task-oriented organizations tend to show corporate proprietorship (i.e. participation in control over property by all members). This is also a primary distinction between the owner-controlled company in a capitalist economy and the giant corporation. In the corporation the machine belongs much more nearly to the worker than to his department chief, much more nearly to the department chief than to the plant manager, and much more nearly to the plant manager than to the company president. This is also, of course, one of the main distinctions between Communist ideology and that of the traditional or capitalist societies it has successfully replaced. It is not, however, a distinction between Communist society and modern Western society.

ORGANIZATIONAL CHANGE

The preceding discussion has suggested or confirmed some ideas as to what elements of organization structure can be appropriately modified in the interest of greater effectiveness in doing work. The pressures toward this kind of change in today's world are of a greater order than organizations have encountered in the past. Thus effective organization more and more requires that adaptation be one of the recognized functions of the permanent executive system.

Basically, many of the problems of adaptation that organizations have been encountering arise from the great reliance they have put on informal organization and intuitive judgment at the highest executive levels. Informal relations are necessarily personal, and they permit the introduction of social considerations that are extraneous to the task to a much greater extent than do procedures that must be stated and defended officially. Moreover, personal confidence and understanding can be built into a large organizational structure only over a long period of time. Changes in personnel disrupt the process and interfere with the informally developed ways of getting things done. Therefore such changes will be resisted in the name of task efficiency as well as presumed injustice to persons.

The result is a need for a system that relies less on informal and more on formal techniques at top levels, especially with respect to the assignment of personnel.

---

[3] Ref. (1), p. 141.

Many practices that have recently come into favor tend to have this effect[4]. First among these is a conscious, planned approach to executive development so that people of appropriate training and proved competence may be available for assignment to key positions as needed. A second is the use of the task-force or project form of organization. This permits the assembly of personnel of a variety of appropriate specialized competences for assignment to the completion of a specific task or project in its entirety. Such groups can be created and broken up rapidly and without a sense of unfairness on the part of the individuals concerned.

Perhaps if we could remove many of the social implications of hierarchical positions, reassignment of key managerial personnel would be a useful device to create rapid and thorough change in larger organization components. If the development of a new and radically revised plan of operation were given to a competent task-force, then that team could be used to replace the previous management, which in turn would be freed for other activities and relieved of the impossible assignment of attempting to learn a new task while still fully occupied in executing an old one. Rotation of personnel in such situations could also be part of a systematic program of executive development.

Increased flexibility in assignment of personnel at the managerial level would also tend toward further professionalization of management. This emphasizes of course the functions of management rather than the hierarchical status. It would imply greater emphasis also on standardized training for managers; more standardized pay scales - and probably lower pay relative to other occupations; mobility, geographically and between organizations; and objective achievement as the basis for remuneration or promotion. At the same time, this tendency would deemphasize policies that rely on organizational loyalty, tenure, hierarchical promotion, and other paternalistic devices.

As with any possibility for action that is openedup by technological progress, organizational change raises questions of value and ideology. One never expects vested interest to approve of divestiture. Likewise, traditions can be defended for their own sake and not judged on the basis of their contribution to material ends. Nevertheless, almost everyone will agree that he should approach a question of choice with as much knowledge as possible. Questions of value can be obscured by sheer ignorance. It is hoped that this paper will contribute a little information to some of these value judgments.

---

[4] Ref. (1), p. 42 ff.

BIBLIOGRAPHY

1. American Society of Mechanical Engineers. 50 Years' Progress in Management, 1910-1960. New York, 1960.
2. Barnard, Chester I. The Functions of the Executive. Cambridge, Massachusetts: Harvard University Press, 1938.
3. Bendix, Reinhard. Max Weber, An Intellectual Portrait. New York: Doubleday & Company, Inc., 1960.
4. Marsh, Robert M. "Formal Organizations in a Pre-industrial Society," Amer. Sociol. Rev. August 1961, Vol. 26, No. 4, p. 547.
5. Mason, Edward S. (Ed.). The Corporation in Modern Society. Cambridge, Massachusetts: Harvard University Press, 1960.
6. Mills, C. Wright. The Power Elite. New York: Oxford University Press, 1956.
7. Moore, W. E. and Feldman, A. Labor Commitment and Social Change in Developing Areas. Social Science Research Council, 1960.
8. Presthus, Robert V. "Weberian v. Welfare Bureaucracy in Traditional Society," Admin. Sci. Quart. June 1961, Vol. 6, No. 1.
9. Schein, Edgar H. "Management Development as a Process of Influence," Industrial Management Review, May 1961, Vol. II, No. 2, pp. 59-77.
10. Selznick, Philip. Leadership in Administration. Evanston, Illinois: Row, Peterson and Company, 1957.
11. Trow, Donald B. "Executive Succession in Small Companies," Admin. Sci. Quart. Sept. 1961, Vol. 6, No. 2.
12. Udy, Stanley H., Jr. "'Bureaucracy' and 'Rationality' in Weber's Organization Theory: An Empirical Study," Amer. Sociol. Rev., December 1959, pp. 791-795.
13. _____, "'Bureaucratic' Elements in Organizations: Some Research Findings," Amer. Sociol. Rev., August 1958, Vol. 23, pp. 415-418.
14. _____, Organization of Work. New Haven, Connecticut: HRAF Press, 1959.

James C. Emery
U.S. Department
of Defense

An Approach to Job Shop Scheduling Using a Large-Scale Computer

I. SCHEDULING

INTRODUCTION

The efficient scheduling of a job shop is an exceedingly complex task. In a large shop thousands of manufacturing orders may be in process at the same time. Each of the orders must be processed through some specified series of operations, and only a limited amount of men and machine capacity exists to perform these operations. As a result, the orders must compete with one another for capacity as they flow through the shop.

The manufacturing order is the document used to authorize the production of a given item. The order may be initiated in response to a specific customer demand or to replenish an inventory account, and it may be for complete apparatus or for only a part or subassembly. Typically the order carries such information as a serial number, a description and identification number of the item to be manufactured, and the quantity to produce. It also may list for each operation, in routing sequence, the work station performing the operation, a description of the operation, the set-up and run time required, and a start time or due date.

Each order is placed in the shop, along with the necessary raw materials, tools, and manufacturing information. Upon the completion of the first operation, the order is transferred to a work station capable of performing the second operation. The order must then wait until capacity becomes available to process it. And so it moves from machine to machine until all of the operations have been completed.

It is the function of scheduling to plan in advance the flow of these orders through the shop and to the customer or storeroom. The goal in scheduling is difficult to define. Should it be to minimize overdue orders? Or idle machine and labor time? Or work-in-process inventory? Or perhaps fluctuations in the manpower level? All of these are desirable objectives, but the attainment of one might be at the sacrifice of the others. Choosing the best objective, or combination of objectives, is extremely

difficult to do in any exact or explicit way. In practice, one must usually settle for a vague compromise among all the goals.

Scheduling of the job shop has a profound effect on a company's competitive position. It is largely through the schedule that management establishes delivery times, manpower plans, work-in-process inventory levels, and capacity utilization. The cost of performing the scheduling function can be quite significant in itself. Because of its extreme complexity and importance, scheduling appears to offer a fruitful field for exploiting the power of large-scale computers.

CONVENTIONAL METHODS OF SCHEDULING

Use of Normal Lead Times. The most commonly used method of scheduling is to date orders on the basis of normal lead times. Normal lead time for an operation includes processing, move and inspection time, and an estimate of the time an order must sit in line waiting to go on a machine.[1] The dating can be done either by working forward from a start date to establish the completion date, or through a process of backdating from the promised shipping date to determine the start dates of assemblies and parts. Forward dating from the start date of the item having the longest total lead time is often used in quoting delivery dates to customers, but detailed scheduling is typically done by the backdating method.

When the backdating method is used, all scheduled dates of assemblies and parts are determined as an automatic consequence of fixing the final shipping date. Once the final shipping date has been set, scheduling the start time of each operation becomes merely a mechanical process of subtracting the operation's lead time from the start time of the next operation. Lead times may sometimes be adjusted to recognize varying load conditions or because of an order's unusual delivery requirements; however, after the lead times have been established, setting an order's final shipping date is the only independent scheduling decision which remains.

Scheduling to Capacity. The other basic method of scheduling is by scheduling to capacity. Under this scheme, available capacity is explicitly considered when orders are scheduled. Operations are scheduled to begin only during a time interval that has not previously been reserved for some other order. When an operation is assigned a start time, the estimated time required is reserved on the machine on which the work will be performed. All operations scheduled subsequently on that machine

---

[1] "Work station" is a more general term than "machine," since some operations are performed without the use of a machine (e.g., hand assembly). However, in order to add concreteness, the following discussion will refer to all work stations as machines.

must then be fitted into some other time interval. If every operation of all orders is scheduled in this manner, the result is a schedule which does not overload the existing capacity.

The most common means of scheduling to capacity is through the use of a Gantt chart. A Gantt chart establishes the sequence of orders processed through a machine and allocates capacity on the basis of an operation's predicted setup and run times. In most cases, the Gantt chart is constructed so that the sequence of orders on the machine can easily be rearranged. In this way, the schedule can be altered to recognize changed delivery requirements or to utilize the machine more effectively.

The distinction between the two methods of scheduling is clearcut. In the backdating procedure, lead times and the final shipping date are established by some means, and the start times of all operations are dated accordingly. The capacity requirement (or "load") during any time interval is only the result of the decisions regarding lead times and the shipping dates. The logic used in scheduling to capacity is quite different. In this case capacity is established first. The schedule is then determined so that the load will not exceed the capacity available. The lead times between operations are merely a consequence of the amount of capacity available and the scheduled sequence at each machine.

DEFICIENCIES IN CONVENTIONAL SCHEDULES

Conventional methods of scheduling have proved satisfactory in a wide variety of job shops. Indeed, the backdating procedure or scheduling to capacity (or some combination of both) lies behind all formalized scheduling systems. Even so, it is important to recognize some rather basic deficiencies in conventional methods of scheduling. The most important of these are: (1) schedules are often not feasible; (2) schedules are usually not effective in meeting delivery dates, reducing inventory, and utilizing capacity; and (3) the scheduling process itself takes so much time that the schedule normally does not reflect current conditions. These three deficiencies do not at all imply faulty administration of the scheduling system. In fact, because of the tremendous complexity of the scheduling problem and the huge mass of data involved, these deficiencies are almost unavoidable in any conventional system.

Infeasible Schedules. The backdating procedure implicitly assumes that the shop has infinite capacity to perform an operation at the scheduled time. Each order is dated without detailed consideration of its effect on all other orders. Shipping dates might be established to provide a level assembly load, but this does not necessarily guarantee that manufacturing sections other than the assembly section will have level loads. Consequently, an operation may be scheduled to start at a time when the machine or machine group performing the operation is already overloaded. Such a schedule is obviously infeasible.

Ineffective Schedules. The problem of infeasible schedules can conceivably be eliminated by scheduling to capacity. The chief obstacle to doing this is the great bulk of data which must be manipulated. But even if such a scheme were practical, the resulting schedules, though feasible, would not necessarily be effective. A feasible schedule might still result in many overdue orders, excessive inventory, and low capacity utilization.

The effectiveness of a schedule established by loading to capacity depends in the final analysis on the scheduled sequence of orders at all machines. Any schedule is, after all, nothing more than a series of sequencing decisions. Even in a moderate-sized shop, an astronomical number of unique schedules can be generated by choosing different sequences of orders at the various machines. Choosing a sequence which results in an acceptable schedule calls for consideration of a large number of inter-related factors.

The problem of considering all of the factors relevant to sequencing is so great that most of them must be neglected when sequencing to capacity by conventional means. But with sufficient manual juggling, orders might be sequenced so that the resulting schedule is a reasonably good one. In most cases, however, this task would be so immense as to make it almost prohibitively expensive. And if cost alone does not limit the practicality of this, the time required to determine a satisfactory schedule usually does.

Time Lag in Scheduling. If a schedule is to serve as a detailed plan of action, it must reflect current conditions in the shop and the latest status of orders. One of the most serious indictments of conventional scheduling methods is that they do not generate sufficiently timely schedules. In many cases, an intolerable time lag exists between a significant event and its recognition in the schedule. The longer the lag, the less likely it is that the shop can use the schedule as anything more than a loose guide.

The problems introduced by the scheduling time lag arise both in scheduling new orders and in rescheduling old ones. No schedule, regardless of its original excellence, will always be followed exactly. Machine breakdowns occur, men fail to report to work, dies break, suppliers fail to meet promised delivery dates, customers cancel orders, and the sales department accepts rush orders. Therefore, to use a currently fashionable word, schedules must be "dynamic" to adjust to unexpected events. This means that schedules must be recalculated periodically to recognize both the latest information concerning the status of orders already scheduled and any new orders received since the last schedule was established.

Rescheduling is often a chaotic affair under conventional methods. Orders must be rewritten or repunched to conform to changed conditions.

Worse yet, orders already released to the shop must be chased down and changed. By the time the changes are effected, conditions on which they were based have usually changed too. The result is seldom worth the effort.

DISPATCHING AND ITS RELATION TO SCHEDULING

Inherent in the actual running of a job shop is the problem of deciding which order should be given priority when two or more are waiting to go on the same machine. Unless a large excess capacity exists, constant competition for machine time among orders is inescapable. It is the function of dispatching to resolve these conflicts.

When a machine completes an operation, the dispatcher must assign the next order for processing on that machine. The order which just completed its operation must then be transferred to the location where its next operation is to be performed or, if all operations have been completed, to the shipping department or a storeroom.

The Conflict Between Dispatching and Scheduling. The dispatcher serves as a link between the scheduling group and the shop. He makes priority decisions within the framework of the schedule. In a sense, however, a conflict exists between the scheduling and dispatching functions. If schedules were completely satisfactory, no dispatching decisions would be necessary; the processing sequence at all machines would be fixed in advance by the schedule.[2] In practice, however, on-the-line dispatching is needed to adjust the schedule to make it more acceptable to the shop. The need to make these dispatching decisions stems, of course, from the fact that the schedules given to the shop are not feasible, effective, or timely.

Dispatchers are assigned to the shop in recognition of this. Much of their work involves trying to make an impossible schedule possible. They can combine orders to save setups, shift work from overloaded to underloaded facilities, and make other on-the-spot adjustments to live with the schedule - but not without limit. A price must be paid for not supplying satisfactory schedules in the first place.

---

[2] Even then the dispatching decisions would not really be avoided; they would merely be shifted to the scheduling function. Scheduling under such a system would actually be advanced dispatching. Although this might conceivably eliminate the need for on-the-line dispatching decisions, there would still exist a need for someone to perform the routine chores commonly assigned to dispatchers.

DISADVANTAGES OF THE CONVENTIONAL SCHEDULING-DISPATCHING SYSTEM

The net effectiveness of conventional scheduling really hinges on the effectiveness of a two-stage process. Scheduling is the first stage and dispatching is the second. The ultimate test of this system is its ability to meet delivery promises, maintain a low level of inventory, and utilize capacity efficiently. In most cases, it falls short on all three counts.

The basic weakness in the system lies in the deficiencies of the schedules given to the dispatcher. Because most schedules are not feasible, effective, or timely, the dispatcher must make adjustments to them. The almost inevitable consequences of relying on the dispatcher to remedy an imperfect schedule are relatively poor delivery performance, excessive inventory, and low capacity utilization.

For one thing, the ad hoc schedule adjustments the dispatcher makes are not at all certain to be the best ones. He cannot possibly keep track of changing customer requirements or conditions throughout the shop to know how best to assign order priorities and to shift loads between facilities. The dispatcher can recognize only a few of the more important - or more immediate - factors which should be considered in making dispatching decisions. Then, too, a dispatcher often faces an army of foremen, expeditors, union stewards, and workers, each with his own parochial (and often conflicting) interest. It would be unreasonable to expect the dispatcher always to act with complete objectivity under these conditions.

But even if the dispatcher made perfect decisions, it is often too late for him to do any more than partially correct for unsatisfactory conditions. For example, unless scheduling predicts in advance unusual capacity requirements, it is difficult to take the necessary steps to alleviate the situation. If delays in shipping are unavoidable, with proper scheduling at least customers can be informed of this in advance. Waiting until problems arise in the shop before action is taken can only result in increased manufacturing costs and the creation of customer ill will.

When the dispatcher makes adjustments to the schedule, then to that extent the schedule failed to predict actual shop events. This robs it of much of its usefulness. Ideally, the schedule should be a reliable and up-to-date prediction of events in the shop. Such a prediction is necessary because of the need to plan material and capacity requirements in advance. As compensation for the uncertainty introduced by an unreliable schedule, some "fat" or "slack" must be added to lead times, extra safety stock must be carried in inventory, and surplus capacity must be reserved.

SUMMING UP

Conventional Scheduling. In conventional scheduling systems, orders are dated either by the backdating process or by scheduling to capacity. The backdating process largely ignores capacity restrictions in the shop, and as a result the schedules so determined are infeasible to some degree. Scheduling to capacity eliminates this objection, but leaves the problem of choosing a sequence of orders at all machines which will lead to effective shop performance. This problem is so enormous that many of the important factors influencing performance cannot be considered. Both the backdating method and scheduling to capacity suffer from the additional disadvantage of taking too long, and as a result the schedule usually does not reflect current conditions.

The dispatcher is handed an imperfect schedule and told to meet scheduled dates, utilize facilities efficiently, and keep inventories low. This is a hopeless demand. It is hopeless, first of all, because of the human limitations of the dispatcher. But even if the dispatcher possessed the gift of omniscience, he could still not compensate for a lack of adequate advanced planning. The result is an increased number of late orders, excess inventories, inefficient utilization of capacity, and wide fluctuations in loads.

II. USE OF A LARGE-SCALE COMPUTER IN SCHEDULING

The extent to which imperfect job shop scheduling contributes to poor delivery performance, higher manufacturing costs, increased capital requirements, and lowered shop morale varies among different shops; but it certainly exists to some degree in all of them. The problem of scheduling is just too complex and too massive to expect otherwise. Large-scale computers offer the only hope of gaining any real mastery over the problem.

Until recently, the relatively small capacity of data processing equipment has limited its usefulness in production control applications. However, the phenomenal capabilities of the present large-scale machines greatly extends their ability to cope with the problem of job shop scheduling.

If a computer is to improve scheduling, it must be able to determine a schedule which is (1) feasible; (2) effective in meeting delivery dates, reducing inventory, and utilizing available capacity; and (3) timely. Such a schedule can be used as a detailed guide in making dispatching decisions. Only then does the schedule serve its full purpose.

FEASIBILITY OF COMPUTER SCHEDULES

A feasible schedule must meet two tests. First, it must not overload existing capacity; and second, it must not schedule the start of an operation before the scheduled completion of the previous operation (and, in many cases, not until sufficient inspection and move time has elapsed). To prevent overloading, an estimate must be made of the available capacity of each machine and the setup and run time required to process each operation of every order scheduled. An even more difficult task is keeping track of each order as it is scheduled from operation to operation so that it will not be scheduled into a machine ahead of the time it will be available from the previous operation.

One way to determine the feasibility of a schedule is to test it in the shop. This, of course, has obvious disadvantages. An alternative to this, and one which tests a schedule prior to its actual execution in the shop, is job shop simulation on a computer.

Job Shop Simulation. Simulation of the shop to develop feasible schedules is a fundamental concept in using a computer for scheduling. In simulating the shop, the computer duplicates internally the flow of orders from machine to machine. A running record is kept in simulated time of each machine's current status (whether the machine is idle or occupied; and if occupied, by which order), its scheduled capacity (number of shifts manned), and the orders waiting to go on the machine. The computer then adjusts the current status appropriately when an event occurs in simulated time. For example, when a machine completes an operation on an order, a new order is scheduled into the machine and the old order is transferred to the waiting line of the machine performing its next operation.

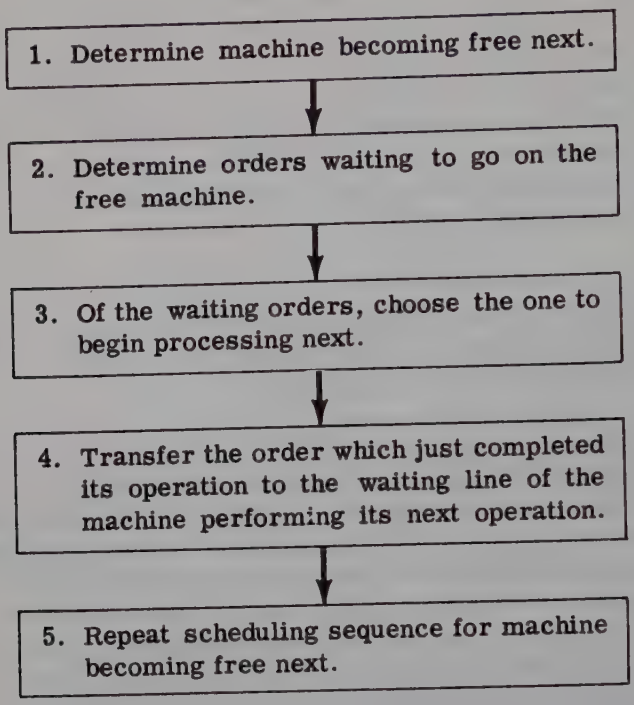
And so the computer moves through time, simulating the movement of orders through the shop. The simulated time at which an operation is begun becomes the scheduled start time for that operation. That is the key. Because the computer is essentially duplicating the shop - but in advance and on a very greatly compressed time scale - the schedule it determines through simulation is entirely feasible if the estimated setup, processing, inspection, and move times are reasonably accurate.

The scheduling cycle which the computer uses can be put in schematic form. The figure on the next page gives a simplified diagram of this cycle. It is basically the same one used by a dispatcher in the shop.

EFFECTIVENESS OF COMPUTER SCHEDULES

Although determining a feasible schedule is by no means a trivial matter, most of its difficulty comes from the huge quantity of data which must be manipulated. The real measure of a schedule is not so much its

feasibility, but its effectiveness in meeting delivery dates, increasing the utilization of available labor and machine capacity, and reducing inventory.


1. Determine machine becoming free next.
2. Determine orders waiting to go on the free machine.
3. Of the waiting orders, choose the one to begin processing next.
4. Transfer the order which just completed its operation to the waiting line of the machine performing its next operation.
5. Repeat scheduling sequence for machine becoming free next.

Computer Scheduling Cycle

The crux of the problem of determining effective schedules lies in Step 3 of the scheduling cycle diagrammed above. The proper selection of the next order to assign to a free machine spells the difference between a good schedule and merely a feasible one. A good schedule can only result from a whole series of good sequencing decisions. The method of making these decisions is through the use of a priority rule.

Priority Rule. A priority rule is used to select the order to assign next to a free machine when two or more orders are competing for the machine. A human dispatcher can get along very easily - if not always efficiently - without detailed priority instructions. A computer cannot. It requires explicit rules for every dispatch "decision" it makes. Since large computers can evaluate extremely complex mathematical expressions with ease, computer scheduling through simulation will permit the use of elaborate priority rules which take many factors into consideration. Some of these factors which a computer might consider are listed below.

1. Due Date. Certainly an order's due date is of prime importance in setting priorities. The nearer the current date approaches an order's due date, the more important it becomes to begin work on that order. The only critical date is the date on which the order is due to be shipped to a customer (or, in the case of an inventory replenishment order, due into the stocking location). However, rather than basing priority decisions at all stages of manufacture on the single shipping date, it would be easier to establish intermediate dates for assemblies and subassemblies. These would then serve as milestones on the path to the final shipping date. The intermediate dates can be established through the backdating procedure. One of the objections to the backdating procedure can partially be eliminated by the computer's ability to make frequent adjustments in lead time in recognition of changes in the load. In any case, the intermediate dates would not be inviolate; they serve merely as convenient guideposts during the simulation and would not be printed out by the computer.

2. Remaining Processing Time. Another important factor in determining priorities is the processing time remaining to be performed on an order. The greater an order's remaining processing time in relation to the time available before it is due, the more critical it becomes to work on the order. Included in the remaining processing time can be estimates of move time, inspection time, heat treating time, and similar necessary operations. In some cases it may even be desirable to estimate the time an order will sit in the waiting lines of the machines performing its remaining operations.

3. Remaining Number of Operations. The total number of operations remaining before an order is completed is often more important than the total remaining processing time. This is especially true for orders having short processing times. Everything else being equal, the order having the largest number of operations remaining before its completion should be given priority over other orders. This factor reflects the inevitable wait time and other unpredictable delays that an order experiences before each operation.[3]

4. Value of Inventory. One method of reducing the value of work-in-process inventory is to schedule high-valued orders on a short lead time basis. This requires that these orders be scheduled to start relatively late. Once started they should then be rushed through the shop. In terms of priorities, the high-valued orders would

[3] If an estimate of wait time is added to the total remaining processing time in determining priorities, it would probably not be necessary to consider the remaining number of operations.

initially be given a low priority in relation to the low-valued orders. When only a minimum amount of slack time remains, they would then be given a high priority so that they will be completed on time.

5. "Criticalness" of Orders. It might be much more important to meet the due dates on some orders than on others. For example, the due dates on defense contracts would certainly be more critical than on parts going into manufacturing stores for which a considerable amount of protective stock is provided. Priorities should reflect this.[4]

6. Possibility of Reducing Setup Cost. When a machine becomes available for work, and one of the orders waiting to go on the machine can be run with little or no additional setup, that order should receive some preference over orders requiring a completely new setup. If setup cost represents a large portion of production cost - as it does, for example, in papermaking and sheet steel slitting - the production sequence which minimizes setup cost may be an overriding consideration in scheduling.

7. Length of an Operation. If, among the orders waiting to be assigned to a machine, the processing time of one of them is considerably shorter than the others, the order with the shortest processing time might be scheduled first. If this were done, the other orders would not be delayed significantly. On the other hand, assigning one of the long jobs to the machine would considerably delay the short job while not appreciably advancing the completion of the long jobs.

8. Effect on Subsequent Machines. Suppose two orders are waiting to be processed on a machine. Suppose also that the machine performing the next operation on one of the orders currently has a long waiting line, while the machine performing the next operation on the other job is presently idle. Other factors being equal, the job next going to the idle machine should receive priority. (Instead of requiring the next machine to be completely idle, a minimum queue length might be used.)

9. Availability of All Parts Going into a Common Assembly. Nothing has yet been said about the problems introduced by an assembly operation where two or more parts are joined together. In order

---

[4]The determination of the "criticalness" of an order, and the weight which should be attached to this factor, would be exceedingly difficult to evaluate in practice. Such a factor would be subject to a certain amount of misuse if management did not employ it with considerable restraint.

to perform such an operation, all of the necessary parts must be available. Therefore, in choosing the assembly order to schedule next, the computer would look only at the orders for which all parts are available. As soon as all of the parts going into an assembly order have been completed in simulated time, the order could then join the assembly waiting line.

10. Relative Progress of Other Parts Going into a Common Assembly. The scheduled progress of all parts going into an assembly should be considered simultaneously in setting priorities. Thus, if the completion of one of the parts for an assembly is unavoidably delayed, the priority of the other parts of the assembly should be reduced so that they will not have to wait around on the assembly floor for the late part. At the same time, the priority of the delayed part should be increased to rush it through its remaining operations.

11. Rescheduling Orders on Idle Machines. In many cases, an operation can be performed on more than one type of machine. For example, a 48-inch lathe can normally perform any operation scheduled on a 36-inch lathe (the converse is not true, of course). Orders are usually routed to the smallest machine capable of performing an operation, for obvious reasons. However, if the larger machine is presently idle, rerouting some order from an overloaded smaller machine may well be justified.[5]

This list of factors is not meant to be exhaustive; it is likely that other factors will have to be considered in calculating practical schedules. A great deal of research is needed to determine the relative importance of each factor and its impact on the over-all effectiveness of a schedule.

Capacity Available for Scheduling. One final point should be mentioned in connection with the determination of feasible and effective schedules. The capacity scheduled by the computer must be fixed over the short term. This, of course, is true in the actual shop; either a machine is manned and available for work, or it is not; and if it is manned, then it is manned for some specific period of time per day and number of days per week. This is not to say the capacity cannot be changed. If an overload exists, additional men can be employed or overtime can be worked. But for any scheduling interval (say an hour or a day or a week), the capacity is fixed, and the schedule should reflect this fact.[6]

[5] Consideration of rescheduling opens a Pandora's box, since it leads into the complex question of the comparative advantage of one machine over another for every operation scheduled. Still another question raised is the possibility of substituting a different labor grade for the "normal" one.

[6] This is an oversimplification. In most shops there are more machines than machine operators. In such a case, the scheduling process must consider the interdependent availability of both men and machines.

In calculating such a schedule, the computer can use a level of capacity fixed in advance by management. Then, if the resulting schedule gives poor delivery performance because of overloads, or low labor utilization and excessive inventory because of underloads, the capacity can be adjusted appropriately and the schedule redetermined.

Such a trial-and-error scheme becomes somewhat cumbersome. A better plan would be to permit the computer to adjust capacities as it moves through simulated time. The computer could first determine the general capacity requirements by looking at the load ahead resulting from standard lead time dating.[7] Then, during the simulation scheduling, the computer could change capacities based on rules which specify the conditions under which a change is economically justified.[8] For example, the computer might be told to schedule an additional two hours overtime if the waiting line behind a machine exceeds, say, ten hours of work, or if orders processed on the machine are more than three days late. If the machine becomes even more overloaded, an additional shift might be added. Similar rules would govern a decrease in capacity when facilities become underloaded or when orders are being delivered too early. The essential point is that the computer can make capacity changes in response to anticipated loads and shifting conditions in simulated time. As part of the output of the computer, a listing would be made of the manpower levels and hours worked for each scheduling interval.

TIMELY COMPUTER SCHEDULES

A timely schedule is one which includes recent events which have significance to the schedule. Timeliness can be achieved in two ways. One way is by a reduction in the number of deviations in the existing schedule. A second way is by a reduction in the time lag between an event and its incorporation in the schedule.

Reducing Scheduling Deviations. A schedule remains timely to the extent that it predicts events accurately. Thus, there is no direct relation between a schedule's timeliness and its age. If the number of scheduling deviations can be reduced, schedules automatically become more durable. Computer scheduling offers a means of eliminating some of the causes of deviations in schedules. It has the added advantage of simplifying rescheduling if this should become necessary.

---

[7] If it is desirable to look ahead further than the load resulting from orders already on the books, forecasts will be necessary. A forecast by representative (or "prototype") models can be used for this purpose.

[8] This would obviously be a difficult job. The cost of changing capacity depends on a complex interaction among such factors as training costs, costs of overtime, costs of inventory, and costs of unemployment compensation.

Deviations in a schedule arise from three basic causes: (1) infeasibility and ineffectiveness of the existing schedule; (2) errors in estimating processing times, and (3) unexpected events, such as added or cancelled orders, engineering changes, absenteeism, and machine breakdown.

1. Infeasibility and Ineffectiveness of the Existing Schedule. One of the primary causes of schedule deviations under present methods is simply that the schedules given to the shop are infeasible and ineffective. To correct for these deficiencies the dispatcher makes adjustments to the schedule. As a result, the schedule fails to give an accurate picture of events in the shop. A computer scheduling system goes at the heart of the problem by reducing these sources of schedule deviations.

2. Errors in Estimating Processing Times. Some variation in processing time is inherent in any production process. The actual time taken to perform a given operation might depend on the particular worker or machine chosen to do the job, the day and hour at which it is done, and a whole series of random events which influence efficiency. Since processing times vary, the estimates of these times will always be subject to some error. The goal is to reduce the variability of the error as much as possible, and to make the estimates "unbiased" in the sense that over the long run errors on the high side are balanced by errors on the low side.

   Computers can minimize the problem of estimating processing times by reducing any bias in the estimate through a detailed analysis of the shop performance. This will help to reduce the number of large deviations in the schedule caused by a gradual accumulation of small errors.

   Lack of bias in estimating processing times obviously does not completely eliminate the problems introduced by an error in estimating the processing time of a particular operation. The only time such an error can have a major effect on the schedule is when an order is not available for assignment into a machine in the sequence specified by the schedule. For example, an order might be somewhat late in arriving at a machine, but if its lateness does not destroy the scheduled sequence of orders on the machine, no permanent damage to the schedule will take place. But once the se-

quence is upset, major deviations in the schedule can occur throughout the shop.[9]

In recognition of the fact that errors exist in the estimates of processing times, the schedule can allow some slack between operations so as to reduce the risk that an order will not be available when scheduled into a machine. An increase in the amount of slack time reduces the risk of disrupting the schedule, but at the cost of increasing inventory.[10]

3. Unexpected Events Causing Major Deviations in the Schedule. By providing the shop with feasible and effective schedules, reducing systematic biases in estimating processing time values, and by allowing sufficient slack time between operations, much of the need for rescheduling can be eliminated. But there will still be occasions, certainly, when rescheduling becomes necessary.

   A computer scheduling system should make rescheduling a relatively simple matter. If the computer is able to schedule economically balanced loads throughout the shop, there will be no need to release an order to the shop more than a short time before its scheduled start time. Rescheduling under these conditions is largely confined to orders not yet released. A change in the schedule will therefore result in very little hardship in the shop: a foreman will not be irked by the reshuffling of an order which he does not know exists. Such changes would involve in many cases no

---

[9] The subsequent effects of disrupting the sequence depend on the stability of a schedule. If a schedule is stable, minor disruptions tend to dampen out. In the case of an unstable schedule, however, even a minor disruption can cause major deviations in the schedule at subsequent machines. The stability which a schedule exhibits is determined by a complex interaction among such factors as the amount of slack the schedule contains, the variability in processing times, the load existing throughout the shop, and the general nature of order routings.

[10] The development of digital controlled machine tools promises to have a major impact on scheduling. With the widespread use of these machines, setup time and run time will be very much more predictable. Setup on such machines often involves very little more than placing punched cards or tape in a sensing device which feeds control data to a machine tool. The actual run time will be machine paced, and therefore quite predictable. The experience gained to date indicates that these machines will have relatively little downtime due to mechanical failures. The combination of high predictability of processing times and low downtime will make it possible to reduce the amount of slack time between operations without adding to the risk of disrupting the schedule. The low setup time will have the additional effect of justifying smaller lot sizes for stocked parts. Reduced lead times and smaller lot sizes will in turn lower inventory requirements. On the other hand, the number of restocking orders - and hence the problem of production control - will grow, and good scheduling will take on added importance.

more than the erasing of magnetic spots in the computer's memory.[11]

Reducing the Time Lag in Calculating a Schedule. The second factor affecting the timeliness of schedules is the time lag taken to include a significant event in the schedule. This time lag has two components. One of these is the time required to calculate the schedule once the necessary data have been collected. This should present no great problem for a computer: speed is a computer's stock in trade. Despite the huge bulk of data which a computer must handle, the time required to calculate a schedule will be quite short.

The other component of the time lag is the time required to collect the data and transform them into a form usable by the computer. This is often the longer of the two time components. The key to exploiting the incredible speed of the computer lies in the rapid collection of data affecting the schedule. Even now devices exist for automatically recording a machine's production and transmitting these data to a central production control office. The real restriction to the widespread application of such devices has been the inability of conventional scheduling systems to utilize rapid feedback information. Computer scheduling does not suffer from such a limitation. When coupled with a computer system, accurate and timely feedback information takes on immense - almost critical - importance.

A COMPUTER SCHEDULING SYSTEM IN OPERATION

A well-conceived computer scheduling system appears to meet all of the requirements of good scheduling. The next question, then, is how can computer scheduling be incorporated into the over-all production control system?

Ultimately the computer system will probably be the production control system. All events in the shop and any changes in orders will automatically and "instantaneously" be transmitted to the computer. Dispatch decisions will be made by the computer in "real time" or "on line." That is, as a machine actually completes an operation and is waiting for work, the computer will select the order to work on next. It will transmit this information directly to the work station or to the foreman in charge of the manufacturing section. In making priority decisions, the computer will consider all of the latest information at its disposal. It will look ahead

---

[11] The foreman would certainly be told in advance the aggregate requirements for manpower, machine capacity, material, and tools. It would not be necessary to give him the detailed composition of the load, however. The only restriction on rescheduling would be that it could not significantly alter the advanced estimates of the aggregate requirements.

by simulating the shop in order to determine the effect of its current decision on future events.

The results of the computer's peering into the future can be printed out periodically as the master schedule. This document would be the basic planning guide. Information appropriate to such groups as line foremen, inventory control personnel, and manpower planners could be extracted from the master schedule and printed in a form most meaningful to the intended users. The computer would revise the schedule only when significant changes have occurred. The time range of the schedule (that is, the length of time it looks ahead) would depend on the amount of advance notice required to hire and train manpower, procure raw materials, add to machine capacity, and so forth.

Such a complete system obviously is quite a few years in the future. Computer scheduling systems will go through several stages before reaching this "ultimate" goal. Probably an early stage will be the periodic - weekly perhaps - calculation of a schedule through the simulation process. In order to limit the amount of computer time required, the detailed schedule might look only a relatively short time into the future. The schedule for events further into the future could be determined by the backdating procedure. Dispatchers in the shop would use the detailed schedule, and would make appropriate adjustments between weekly calculations.

As computers become more powerful and greater knowledge is gained in their use, more and more of the production control routine could be shifted to the computer. The frequency of calculating schedules could be increased, the range of schedules could be lengthened, and more rapid and automatic feedback devices could be incorporated into the system. The result of this continual process of improvement and extension will almost surely be a real time computerized production control system.

LIMITATIONS TO COMPUTER SCHEDULING

The discussion so far has pointed out only the advantages and none of the limitations of computer scheduling. The picture is not all that rosy, however. There are some important limitations to the application of computers in job shop scheduling. These limitations are of either a technical or economic nature.

Technical Limitations. Priority scheduling of even a moderately-sized shop involves a staggering amount of data shuffling. A computer capable of manipulating the data rapidly enough must possess ravenous input capabilities, an exceedingly fast arithmetic unit, and a huge high-speed random access memory.

As severe as these requirements are, some of the largest machines now available or on the immediate horizon appear to offer the necessary characteristics. The rapid strides being made in computer technology certainly indicate that if presently available hardware does impose any limitations, these limitations will be short-lived. Each year sees a many-fold increase in computer capacity, and there is no reason to believe that this trend will abate.

Perhaps a more fundamental technical limitation is the programmer's inability fully to tap the power which the computer engineer designed into his machine. A number of extremely difficult problems must be faced and solved before the first satisfactory simulation scheduling system is installed. Not the least of these is the problem of determining an effective priority rule. An almost human logical ability must be breathed into the computer before it can make dispatching decisions which result in uniformly good schedules.

But surely this problem is no more difficult than some of the other complex tasks which computers have tackled with success. Computers have been programmed to play a good game of checkers and a tolerable game of chess, to prove difficult logical and geometric theorems, and to read ill-defined shapes and patterns. The effort that has gone into developing these programs is not primarily for amusement; instead, it is meant to demonstrate the powerful logical ability of computers. The current work directed toward developing increasingly sophisticated computer languages will even magnify this ability.

Economic Limitations. Economic considerations, rather than technical limitations, place the greatest restrictions on the use of computers for job shop scheduling. The cost of large scale computers has been soaring. However, this increase has been at a much slower pace than the increase in capacity. The result is that the cost of performing a given series of calculations on today's computers is very much less than the cost on the best machine available only a few years ago. The extensive use of large-scale machines and the recent developments in computer technology seem destined to make the computations involved in job shop scheduling entirely feasible from an economic standpoint.

The expense of the hardware itself is only one of the costs of developing a system of computer scheduling. There is simply no denying that an enormous amount of effort must go into developing any computer system, and the complexity of the job shop scheduling program makes it an especially expensive application. Skilled programmers and systems experts must be employed, new procedures must be set up, and all personnel, from manager to shop worker, must be trained in the use of the new system.

All of this costs money. But so do low machine utilization, excessive inventory, large fluctuations in work loads, squadrons of expeditors, and customer ill will generated by poor delivery performance. As plants become more automated, many of these costs will increase. The question is, which is more expensive, computer control or poor scheduling? Certainly a manager must face the fact that a computer scheduling system will be exceedingly expensive to install. On the other hand, he should also recognize that not developing such a system might entail the even more unpalatable alternatives of lower efficiencies, higher inventories, and a gradual erosion of competitive position.

CONCLUSION

Within the foreseeable future - say within the next ten years - routine scheduling of a job shop on computers will almost certainly exist in a number of progressive companies. There are two excellent reasons for confidence in this: (1) the need for such a system is very great and (2) there appears to be no permanent technological or economic block to developing such a system. Anyone even vaguely familiar with a large job shop recognizes the need. The only possible hope for obtaining any real degree of control over such a shop is through the use of a large-scale computer. The equipment necessary to do the job is already at hand or in sight; the real problem lies in developing the system to use the available hardware. Priority scheduling through job shop simulation appears to offer the most promising approach to solving this problem.

REFERENCES

Conway, R. W., Johnson, B. M. and Maxwell, W. L. "A Queue Network Simulator for the IBM 650 and Burroughs 220," Communications of ACM, December, 1959.

Conway, R. W., Johnson, B. M. and Maxwell, W. L. "An Experimental Investigation of Priority Dispatching," The Journal of Industrial Engineering, May-June 1960.

Conway, R. W., Johnson, B. M. and Maxwell, W. L. "Some Problems of Digital Systems Simulation," Management Science, October 1959.

Holt, Modigliani and Simon. "Linear Decision Rule for Production and Employment Scheduling," Management Science, October 1955 and January 1956.

Rowe, A. J. "Sequential Decision Rules in Production Scheduling," Doctoral Thesis, UCLA, August 1958.

Sesson, Roger L. "Methods of Sequencing in Job Shops - a Review," Operations Research, January-February 1959.